

Erikson Kaszubowski

## **Modelo de tópicos para associações livres**

Tese apresentada como requisito parcial para a obtenção do grau de Doutor em Psicologia pelo Programa de Pós-Graduação em Psicologia, Centro de Filosofia e Ciências Humanas da Universidade Federal de Santa Catarina.

Orientador: Prof. Dr. Fernando Aguiar Brito de Sousa

Florianópolis  
2016

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Kaszubowski, Erikson

Modelo de tópicos para associações livres / Erikson  
Kaszubowski ; orientador, Fernando Aguiar Brito de Sousa -  
Florianópolis, SC, 2016.

227 p.

Tese (doutorado) - Universidade Federal de Santa  
Catarina, Centro de Filosofia e Ciências Humanas. Programa  
de Pós-Graduação em Psicologia.

Inclui referências

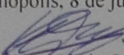
1. Psicologia. 2. Psicanálise. 3. Modelos de tópicos. 4.  
Associações livres. 5. Modelagem probabilística. I. Sousa,  
Fernando Aguiar Brito de. II. Universidade Federal de  
Santa Catarina. Programa de Pós-Graduação em Psicologia.  
III. Título.

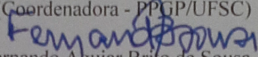
*Erikson Kaszubowski*

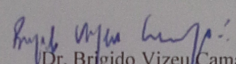
*Modelo de tópicos para associações livres*

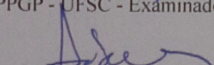
Tese aprovada como requisito parcial à obtenção do grau de Doutor em Psicologia, Programa de Pós-Graduação em Psicologia, Centro de Filosofia e Ciências Humanas da Universidade Federal de Santa Catarina.

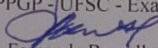
Florianópolis, 8 de julho de 2016

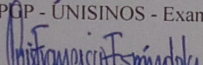
  
Dra. Carmen Leontina Ojeda Ocampo Moré  
(Coordenadora - PPGP/UFSC)

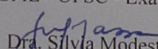
  
Dr. Fernando Aguiar Brito de Sousa  
(PPGP - UFSC - Orientador)

  
Dr. Brígido Vizeu Camargo  
(PPGP - UFSC - Examinador)

  
Dr. Roberto Moraes Cruz  
(PPGP - UFSC - Examinador)

  
Dra. Fernanda Barcellos Serralta  
(PPGP - UNISINOS - Examinadora)

  
Dr. Luis Francisco Espíndola Camargo  
(PPGFIL - UFSC - Examinador)

  
Dra. Silvia Modesto Nassar  
(PPGCC - UFSC - Examinadora)

Dra. Daniela Ribeiro Schneider  
(PPGP - UFSC - Suplente)

Dr. Adriano Beiras  
(PPGP - UFSC - Suplente)

Prof. Dr. Carlos Henrique Sancelmo da Silva Nunes  
Programa de Pós-Graduação em Psicologia - CFH/UFSC  
Coordenador - Portaria Nº 898/2015/GR



## Resumo

Apesar da reiterada importância da experiência clínica na elaboração, avaliação e reformulação das teorias psicanalíticas, a maioria das pesquisas em Psicanálise no Brasil ou se afasta da clínica completamente ou faz uso esparso e questionável de seu material. Para estabelecer um elo mais rigoroso entre os dados da prática da Psicanálise e suas teorias, propomos a elaboração de um modelo formal das associações livres. O modelo proposto é fundamentado, por um lado, nos modelos elaborados por Freud sobre o mecanismo de funcionamento das associações livres, e, por outro, nos modelos probabilísticos de tópicos, como a Alocação de Dirichlet Latente. Para avaliar o modelo, conduzimos um estudo de caso cujo *corpus* é baseado na gravação em áudio e transcrição integral de 47 sessões de psicanálise. Após tratamento do *corpus*, ajustamos o modelo final por meio do algoritmo do *Structural Topic Model*, considerando os turnos de fala como unidade de contexto, propondo a identificação de 50 tópicos e utilizando o número da sessão e a identificação do interlocutor como covariáveis. O modelo ajustado foi validado externamente por meio do uso de estimativas dos parâmetros na predição de variáveis de critério e na predição dos resultados de uma bateria de testes de associação de palavras. Modelos de *random forests* que utilizaram o parâmetro de proporções de tópicos estimado pelo modelo como preditor foram capazes de distinguir, com pouco erro em validação cruzada, o interlocutor e o número da sessão aos quais pertencem os turnos de fala. No teste de associação de palavras, utilizando a distribuição marginal dos termos como referência, o modelo proposto fez predições superiores para 63% a 65% dos pares de palavras, quando consideramos, respectivamente, a probabilidade e o ranque preditos para as palavras associadas. Por fim, para demonstrar a utilidade do modelo em produzir informações relevantes sobre o caso, apresentamos inferências em termos do conteúdo dos principais complexos identificados automaticamente pelo modelo, sua correlação na formação de agregados, sua prevalência nos turnos de fala do analista e do analisando e sua dinâmica temporal ao longo do tratamento.

**Palavras-chave:** associações livres, modelos de tópicos, estudo de caso, psicanálise, formalização

## Abstract

Clinical evidence has an important place in the elaboration, evaluation, and reformulation of psychoanalytic theory. But most Brazilian research in Psychoanalysis either ignores clinical evidence completely or uses it in a sparse and questionable manner. We propose a formal model for free association as a mean to build rigorous relations between clinical evidence and psychoanalytic theory. Our formal model is based, on one side, on Freud's models about free association operating mechanisms, and on probabilistic topic models, such as Latent Dirichlet Allocation, on the other side. We evaluate the proposed model with a clinical case study based on the audio recording and full transcription of 47 psychoanalytic sessions. After cleaning up the *corpus*, the final model was adjusted using the *Structural Topic Model* algorithm, using each turn of speech as a document, defining 50 topics and predicting topic proportions by means of session number and speaker identity. The model validation was based on two prediction tasks: predicting criterion variables using parameter estimates and predicting the results of word-association tests. *Random forest* models that used topic proportion parameters as predictor were able to distinguish, with low cross-validation error, the speaker and session number of the turns of speech. For the word-association tests, the topic model made superior predictions for associated word probability and rank for 63% and 65% word pairs, respectively, when compared to a baseline model based on marginal word type distribution. We present model-based inferences in order to demonstrate the proposed model's ability to produce relevant information about the case study. We focus on the content of the complexes automatically identified by the model, their correlation and graph communities, prevalence on each speaker turn of speech and temporal dynamics during treatment.

**Keywords:** free association, topic models, case study, psychoanalysis, formalization

## Résumé

Malgré l'importance répétée de l'expérience clinique dans la préparation, l'évaluation et la reformulation des théories psychanalytiques, la plupart des recherches en psychanalyse au Brésil ou se écarte de la clinique ou ne utilise son matériel que rare et douteusement. Pour établir un lien plus rigoureuse entre des données de la clinique et des théories psychanalytiques, nous proposons l'élaboration d'un modèle formel des associations libres. Le modèle proposé est basé, d'une part, sur les modèles développés par Freud sur le mécanisme de fonctionnement des associations libres, et, deuxièmement, sur les modèles probabilistes de sujets, tels que l'Allocation de Dirichlet Latente. Pour évaluer le modèle, nous avons conduit une étude de cas dont le *corpus* est basé sur l'enregistrement audio et transcription complète de 47 sessions de psychanalyse. Après le traitement du *corpus*, nous avons ajusté le modèle final à travers de l'algorithme *Structural Topic Model*, en utilisant des tour de parole comme unité de contexte, proposant l'identification des 50 sujets et en utilisant le numéro de session et l'identification de l'interlocuteur comme co-variables. Le modèle ajusté a été validé externment en utilisant des estimations des paramètres pour la prédiction de variables de critère et en prédisant les résultats d'une batterie de tests d'association de mots. Modèles de *random forests* qui ont utilisé le paramètre de proportion de sujets comme un predicteur ont été en mesure de faire la distinction, avec peu d'erreur en validation croisée, de l'interlocuteur e du numéro de la session à laquelle appartiennent les tours de parole. Dans les tests d'association de mots, en utilisant de la distribution marginale des termes comme référence, le modèle proposé a faire des prédictions supérieures pour 63% à 65% des paires de mots, si l'on considère, respectivement, la probabilité et le rang prédit pour les mots associés. Enfin, pour démontrer l'utilité du modèle pour produire des informations pertinentes sur le cas, nous présentons des inférences sur le contenu des complexes identifié automatiquement pour le modèle, la corrélation et la formation d'agrégats, sa prévalence dans les tour de parole de l'analyste et de l'analysant, et la dynamique temporelle pendant le traitement.

**Mots-clés:** associations libres, modèle de sujet, étude de cas, psychanalyse, formalisation





# Lista de Figuras

2.1	Exemplo de modelo gráfico probabilístico . . . . .	56
2.2	Modelo gráfico probabilístico da LDA . . . . .	67
2.3	Modelo gráfico probabilístico do unigrama . . . . .	69
2.4	Modelo gráfico probabilístico para a mistura de uni-gramas . . . . .	69
2.5	Modelo gráfico probabilístico da pLSI . . . . .	70
3.1	Modelo de tópicos para as associações livres . . . . .	99
3.2	Modelo gráfico probabilístico do STM . . . . .	102
4.1	Gráfico de linha do logaritmo da contagem de pala- vras em cada turno de fala do <i>corpus</i> não tratado . .	114
4.2	Gráfico da função de probabilidade cumulativa em- pírica da contagem de palavras por turno de fala do <i>corpus</i> não tratado . . . . .	115
4.3	Gráfico de Zipf do <i>corpus</i> não tratado . . . . .	116
4.4	Gráfico de linha do logaritmo da contagem de pala- vras em cada turno de fala para o <i>corpus</i> processado	124
4.5	Gráfico da função de probabilidade cumulativa em- pírica da contagem de palavras por turno de fala no <i>corpus</i> tratado . . . . .	125
4.6	Gráfico de Zipf do <i>corpus</i> tratado. . . . .	126
4.7	Verossimilhança de dados retidos, resíduos, exclusi- vidade e coerência semântica do modelo ajustado em função do número de tópicos. . . . .	129
4.8	Limite inferior da evidência para seis modelos de 50 tópicos . . . . .	133
4.9	Distribuição dos escores de desvio dos tópicos. . . . .	136
4.10	Checagem Preditiva Posterior da Informação Mútua para os dez primeiros termos de três tópicos . . . . .	137
4.11	Distribuição da razão de probabilidade dos pares as- sociativos . . . . .	147
5.1	Gráfico das características do complexo 21. . . . .	160
5.2	Gráfico das características do complexo 35 . . . . .	162

5.3	Gráfico das características do complexo 31 . . . . .	165
5.4	Gráfico das características do complexo 46 . . . . .	168
5.5	Gráfico das características do complexo 11 . . . . .	169
5.6	Gráfico das características do complexo 43 . . . . .	173
5.7	Gráfico dos termos dos complexos 18 e 12 . . . . .	175
5.8	Gráfico das características do complexo 37 . . . . .	179
5.9	Gráfico das características do complexo 26 . . . . .	180
5.10	Gráfico das características do complexo 3 . . . . .	183
5.11	Gráfico das características do complexo 30 . . . . .	186
5.12	Gráfico das características do complexo 17 . . . . .	188
5.13	Probabilidade dos complexos . . . . .	191
5.14	Grafo de correlações entre complexos . . . . .	192

# Lista de Tabelas

4.1	Informações descritivas do <i>corpus</i> não tratado. . . .	113
4.2	Frequência absoluta e relativa das 47 primeiras palavras do <i>corpus</i> não tratado . . . . .	117
4.3	Colocações baseadas em trigramas e bigramas . . . .	119
4.4	Informações descritivas do <i>corpus</i> tratado. . . . .	123
4.5	Frequência absoluta e relativa das 20 primeiras palavras do <i>corpus</i> tratado. . . . .	127
4.6	Pares associativos mais comuns . . . . .	146
4.7	Ranque inverso das palavras associadas sob o modelo de tópicos e o modelo marginal . . . . .	150



# Sumário

<b>Introdução</b>	<b>1</b>
Questionando o lugar da pesquisa em Psicanálise . . . . .	1
Modelos analógicos e digitais . . . . .	7
Por um modelo digital das associações livres . . . . .	11
Escopo da tese . . . . .	14
<b>1 Associações livres</b>	<b>17</b>
1.1 Nossos antecedentes . . . . .	17
1.2 Origem do método de associação livre . . . . .	20
1.3 Modelos analógicos apresentados por Freud . . . . .	22
1.3.1 Modelo geológico e modelo têxtil . . . . .	27
1.3.2 Representações-meta e associacionismo . . . . .	29
1.3.3 O conceito de complexo . . . . .	30
1.3.4 A necessidade de vários registros do psiquismo . . . . .	32
<b>2 Modelos de tópicos</b>	<b>35</b>
2.1 Inferência bayesiana . . . . .	36
2.1.1 Regra de Bayes . . . . .	37
2.1.2 Regra de Bayes aplicada à testagem de hipóteses . . . . .	39
2.1.3 Estimativa de parâmetros . . . . .	40
2.1.4 Computando a distribuição posterior . . . . .	45
2.2 Modelos gráficos . . . . .	53
2.2.1 Redes bayesianas . . . . .	54
2.2.2 Dependência, independência e independência condicional . . . . .	56
2.2.3 Modelos gráficos como modelos gerativos . . . . .	58
2.3 Alocação de Dirichlet Latente . . . . .	60
2.3.1 Problema motivador e antecedentes . . . . .	60
2.3.2 Notação e terminologia . . . . .	62
2.3.3 A LDA como modelo gerativo . . . . .	64
2.3.4 A LDA como modelo gráfico . . . . .	66
2.3.5 A LDA como modelo probabilístico . . . . .	71
2.3.6 Estimativas dos parâmetros da LDA . . . . .	73
2.3.7 Desenvolvimento de outros modelos de tópicos . . . . .	74

2.4	Utilização dos modelos de tópicos na Psicologia . . .	78
2.4.1	O uso de modelos de tópicos para análise de associações livres . . . . .	82
<b>3</b>	<b>Modelos e procedimentos</b>	<b>85</b>
3.1	Objeto modelo das associações livres . . . . .	86
3.1.1	Predicados de caracterização . . . . .	87
3.1.2	Predicados de delimitação . . . . .	89
3.1.3	Predicados de determinação . . . . .	90
3.1.4	Definição formal . . . . .	92
3.2	Modelo digital das associações livres . . . . .	93
3.2.1	Dos complexos associativos . . . . .	94
3.2.2	Das representações-meta . . . . .	95
3.2.3	Das associações livres . . . . .	96
3.2.4	Síntese do modelo digital . . . . .	97
3.3	Implementação do modelo: o pacote <i>stm</i> . . . . .	98
3.3.1	Diferenças com o modelo proposto . . . . .	103
3.4	Procedimentos do estudo de caso . . . . .	105
3.4.1	Protocolo e contato inicial . . . . .	106
3.4.2	Condução das sessões . . . . .	107
3.4.3	Transcrição . . . . .	108
3.4.4	Teste de associação de palavras . . . . .	108
3.4.5	Local de realização da pesquisa e considerações éticas . . . . .	109
<b>4</b>	<b>Seleção e validação do modelo</b>	<b>111</b>
4.1	Descrição do <i>corpus</i> . . . . .	112
4.2	Tratamento dos dados . . . . .	118
4.2.1	Identificação de colocações . . . . .	118
4.2.2	Lematização . . . . .	120
4.2.3	Remoção de <i>stopwords</i> . . . . .	121
4.2.4	Descrição do <i>corpus</i> tratado . . . . .	122
4.3	Seleção do modelo . . . . .	126
4.3.1	Número de tópicos . . . . .	127
4.3.2	O problema da multimodalidade . . . . .	130
4.3.3	Decisão final . . . . .	132
4.4	Análise do ajuste aos pressupostos . . . . .	133
4.5	Validação do modelo por predição de dados de teste	138
4.6	Validação do modelo por teste de associação de palavras	140
4.6.1	Modelando o teste de associação de palavras	141
4.6.2	Hipóteses . . . . .	144

---

4.6.3	Resultados . . . . .	145
4.6.4	Sensibilidade dos resultados à escolha do modelo	150
4.6.5	Considerações . . . . .	151
<b>5</b>	<b>O caso Ivan Ivanovich</b>	<b>153</b>
5.1	Apresentação . . . . .	154
5.2	Principais complexos . . . . .	157
5.2.1	Término da dissertação . . . . .	159
5.2.2	Retorno a Sevastopol . . . . .	164
5.2.3	A mãe e o lar . . . . .	168
5.2.4	Raiva e mal-estar . . . . .	172
5.2.5	Nina e as mulheres . . . . .	177
5.2.6	Memórias do avô . . . . .	182
5.2.7	<i>Wanderlust</i> . . . . .	184
5.2.8	Barreiras à associação livre . . . . .	187
5.3	Outras inferências do modelo . . . . .	189
5.3.1	Probabilidade dos complexos . . . . .	190
5.3.2	Rede de proximidade entre complexos . . . . .	191
5.4	Considerações . . . . .	193
	<b>Conclusão</b>	<b>195</b>
	Limitações . . . . .	196
	Aplicações e perspectivas . . . . .	198
	Momento de concluir . . . . .	200
	<b>Referências</b>	<b>201</b>
<b>A</b>	<b>Código-fonte e visualização</b>	<b>213</b>





# Introdução

## Questionando o lugar da pesquisa em Psicanálise

Quando confrontado sobre o lugar que ocupa a pesquisa na Psicanálise, um psicanalista pode calmamente responder, parafraseando as palavras célebres de Freud (1984g), que pesquisa e tratamento coincidem. Autorizado pelo próprio texto freudiano, afirma que pesquisa em Psicanálise implica necessariamente na condução de atendimentos clínicos.

Sem dúvida, a clínica teve e ainda tem, na constituição dos andaimes da teoria psicanalítica, um lugar fundamental que permitiu a Freud (1984b) sempre demarcar com clareza que, se podemos falar em uma metapsicologia, não devemos confundir-la em nenhum momento com uma *Weltanschauung* – a não ser a própria visão de mundo da ciência – uma vez que a teoria é apenas o andar superior de um fundamento calcado nas evidências do trabalho clínico.

Mesmo assim, a facilidade com que as palavras de Freud são reafirmadas contrasta com o que de fato é praticado. De que maneira o psicanalista trata a evidência que surge na clínica? Qual a natureza dessa evidência? Como ela é comunicada? As pesquisas que tratam de casos clínicos não passam, na maioria das vezes, de vinhetas retiradas das sessões ao sabor das intenções do pesquisador ou apanhados gerais cuja fidedignidade pode ser sempre questionada, pois não há meios para avaliar os vieses introduzidos pelo psicanalista, contrastando os estudos de caso clínicos com os estudos de caso empíricos, necessariamente mais rigorosos, mas relativamente mais raros na literatura psicanalítica (Serralta, Nunes, & Eizirik, 2011). Pelo contrário, os vieses são interpretados como componentes legítimos do processo transferencial da análise, e a justificativa de que eles não interferem na fidelidade do relato se assenta na qualidade da análise pela qual passou o pesquisador (Guimarães & Bento, 2008). Ou, por outro lado, o caso é descrito a partir de seu encaixe retroativo nos conceitos psicanalíticos (Mezan, 1998), pouco

servindo como evidência da validade desses conceitos, portanto.

Ainda que a clínica seja exaltada como principal fonte de evidência para qualquer fundamentação da teoria psicanalítica, são poucos e raros os casos clínicos descritos de maneira mais integral, tal como empreendido por Freud. Relatos de casos que se aproveitaram dos avanços tecnológicos para registro das sessões em mídia analógica ou digital são exceção, como o caso de Mrs. C, de Dahl, e Amalia X, de Thomä e Kächele (Kächele, Schachter, & Thomä, 2011). Esses e outros casos foram reunidos no *Ulm Textbank* (Dahl, Kächele, & Thomä, 1988), uma iniciativa do Instituto Psicanalítico de Ulm, na Alemanha, preocupada com a investigação empírica das conjecturas psicanalíticas, mas que hoje é de acesso fechado e pago. É claro que as preocupações éticas relativas à identificação dos pacientes estão mais prementes agora e limitam os detalhes dos estudos de caso, mas a experiência do *Ulm Textbank* deixa evidente que é injustificado simplificar um caso clínico a um número reduzido de excertos selecionados intencionalmente, como também é injustificada a recusa em utilizar ferramentas que permitam um registro mais confiável das sessões.

A preocupação com os efeitos da subjetividade do psicanalista na seleção enviesada de excertos de sua prática clínica ecoa um problema antigo para a Psicanálise. Trata-se do motivo para rompimento da relação entre Freud e Fliess depois de seu último congresso em Achensee, no verão de 1900. Ainda que outros eventos possam ter contribuído para o desgaste da amizade entre os dois, o ataque de Fliess à maneira como Freud empreendia sua investigação dos fenômenos psíquicos se mostrou um golpe de misericórdia. Fliess argumentou que a interpretação que Freud elaborava para os sintomas neuróticos e outros derivados do inconsciente não passava de uma projeção de sua própria mente, pois as modificações no psiquismo não poderiam ser creditadas à Psicanálise, mas à periodicidade dos fenômenos psíquicos (Freud, 1986). O impacto do golpe foi tal que, mesmo um ano mais tarde, numa das cada vez mais raras correspondências, Freud responde dizendo que a acusação de ser um “leitor de pensamentos” colocava em cheque todo o valor de sua pesquisa (Carta 270, Freud, 1986).

Essas dificuldades retornam principalmente na pesquisa psicanalítica no contexto universitário. Se algumas instituições de Psicanálise podem aceitar excertos clínicos e a experiência pessoal como prova suficiente para a teoria e prática psicanalíticas, a pesquisa universitária exige uma melhor demarcação do caráter científico das

pesquisas empreendidas. O que pode responder um psicanalista a respeito de sua pesquisa quando ela está inserida num contexto acadêmico? Ou ainda: poderia a Psicanálise vir a constituir métodos de pesquisa aceitáveis no contexto universitário? A pergunta não é nova, e temos várias respostas propostas para ela.

Podemos iniciar citando o trabalho já clássico empreendido por Jean Laplanche (1980) na Universidade de Paris VII: a aplicação de determinadas categorias do método clínico psicanalítico à leitura do próprio texto que compõe a teoria da Psicanálise. Esse tipo de pesquisa histórico-conceitual é esposado por Luiz Alfredo Garcia-Roza (1994) em sua proposta de pesquisa psicanalítica na universidade. Caberia ao pesquisador psicanalista empreender, no contexto universitário, uma releitura da teoria psicanalítica, aproveitando-se da textualidade do corpo teórico para investigar “a potência significante enquanto gerador de múltiplos sentidos” (Garcia-Roza, 1994, p. 15). As dificuldades em lidar com os dados clínicos, especialmente no contexto universitário, acabariam impondo como solução a realização de trabalhos teóricos de releitura da Psicanálise.

Tal posição, porém, limita o escopo das pesquisas em Psicanálise na universidade à interpretação de um *corpus* de obras clássicas, tal como comumente empreendido na Filosofia. Essa alternativa não é a única, contudo, e é relativizada por Renato Mezan (2002) em suas discussões sobre a pesquisa psicanalítica na universidade, no sentido de legitimar práticas diversas que escapam da proposta histórico-conceitual. Para Mezan, há pelo menos mais dois tipos de pesquisa psicanalítica além da proposta de releitura teórica: aquela que parte de material clínico e a Psicanálise aplicada. Essas propostas se diferenciam tanto no material empírico sobre o qual operam quanto com relação aos métodos utilizados. A pesquisa com material clínico implica na reflexão sobre a experiência psicanalítica no sentido de formalizar e compreender o que se passa com um determinado paciente em seu processo de análise, podendo daí emergir contribuições teóricas originais ou reiteraões de compreensões teóricas já consagradas. A Psicanálise aplicada, por sua vez, debruça-se sobre algum tipo de fenômeno cultural – uma obra de arte, um ritual, sistemas de crenças ou saberes científicos – e utiliza a rede conceitual da Psicanálise para lançar novas luzes sobre esses fenômenos, revelando aspectos inconscientes, transferenciais ou de outro modo ligado à maneira como a Psicanálise concebe o homem e suas produções.

Abordar a questão da pesquisa em Psicanálise na universi-

dade, portanto, implica em considerar seus desdobramentos em múltiplos projetos metodológicos, com objetos e procedimentos de investigação diversos. Essa multiplicidade é, muitas vezes, adotada de maneira radical: não caberia à pesquisa em Psicanálise ater-se a uma regulamentação que previsse cada problema a ser encontrado, no sentido de elencar procedimentos específicos a serem executados a cada caso. Por isso, uma pesquisa em Psicanálise só poderia desenvolver seu método adequando-se às especificidades de seu problema e, portanto, os caminhos metodológicos significam o próprio desdobramento de um problema de pesquisa, reconstruções únicas que tendem a generalizações crescentes (Aguiar, 2006).

A maioria dos estudos, todavia, permanece congelada no primeiro passo. Os métodos são, muitas vezes, desenvolvidos de forma a se adequar aos problemas singulares de cada pesquisa. O recurso de adequação feita ao longo do caminho e a dificuldade em operar de fato generalizações crescentes, porém, indicam um primeiro problema nessa compreensão radical de como as pesquisas devem ser empreendidas. Um dos motivos da explosão metodológica desorganizada, sem convergência possível, é que as formulações teóricas utilizadas na Psicanálise são genéricas demais para poderem prever um resultado a ser posto à prova por uma pesquisa bem delimitada. A maioria das pesquisas universitárias em Psicanálise, independentemente de serem histórico-conceituais, clínicas ou aplicadas, contornam essa dificuldade utilizando os conceitos psicanalíticos como chaves hermenêuticas de interpretação, reencontrando nos resultados das pesquisas os próprios conceitos que lhe serviram de fundamento (Grünbaum, 1985). A justificativa desses procedimentos descritivos e interpretativos, alinhados com a cisão entre as *Naturwissenschaften* e as *Geisteswissenschaften* de Dilthey, apela para epistemologias regionais e para a especificidade dos objetos e métodos de estudos da Psicanálise, esquivando-se do problema do valor científico do conhecimento produzido e racionalizando seus problemas sob um regime que permite exceções ao rigor científico (Edelson, 1990).

A circularidade hermenêutica e a capacidade de gerar explicações *a posteriori* para todo e qualquer evento psíquico, a despeito de sua aparência de robustez, levou Popper (1962) a definir a Psicanálise como uma pseudociência. O critério de demarcação proposto por Popper exige que uma ciência deduza, partindo de suas conjecturas teóricas, enunciados arriscados passíveis de serem refutados por meio de observações ou experimentos precisos e bem delimita-

dos. A Psicanálise, porém, careceria de predições arriscadas e construiria, em seu lugar, verificações de suas conjecturas com base em heurísticas não-científicas. Uma dessas heurísticas é a analogia do quebra-cabeça, reiterado por Freud como critério de adequabilidade de uma interpretação ao conteúdo manifesto de um sonho, e baseado no princípio de que uma interpretação poderia ser considerada correta com base em quão bem ela consegue encaixar os diferentes elementos do conteúdo manifesto numa narrativa coerente sobre os pensamentos latentes (Meehl, 1983). Apesar do poder argumentativo persuasivo da analogia do quebra-cabeça, ela só pode ser aceita como um recurso retórico, pois não exclui a possibilidade da existência de outras explicações para o conteúdo manifesto, inclusive aquelas alheias ao escopo da teoria psicanalítica.

O problema da cientificidade é um capítulo à parte que pertence à própria história da Psicanálise, e não nos aprofundaremos nele aqui (para uma breve revisão do assunto, cf. Fontes, 2015). Acrescentaremos apenas que a tese de Popper sobre a Psicanálise é fortemente combatida por Adolf Grünbaum (1985). Para Grünbaum, a Psicanálise oferece, de fato, enunciados derivados de suas conjecturas que poderiam ser refutadas por observações ou mesmo experimentos bem delimitados. Um exemplo é a teoria da origem dos sintomas neuróticos a partir do recalque dos representantes de pulsões conflitantes. Se um sintoma de fato se origina a partir desse processo, e somente a dissolução do recalque pode fazê-lo desaparecer, a resolução do sintoma por outro método ou terapia consistiria em evidência da falsidade da teoria do recalque, pelo menos como originadora de sintomas psíquicos. E, de fato, Freud toma o desaparecimento de sintomas específicos a partir da recordação de memórias recalçadas como corroboração de suas teorias. O problema, porém, é que sintomas podem entrar em remissão espontaneamente, e outros métodos psicoterapêuticos também demonstram eficácia na redução ou desaparecimento de sintomas neuróticos (Grünbaum, 1977) – como sintetizar as evidências que apontam para direções diferentes? O recurso retórico empregado pelos psicanalistas consiste em reduzir o efeito de tratamentos alternativos à mera substituição de sintomas, de forma a evitar revisar a teoria, mesmo não havendo nenhuma evidência conclusiva dessa substituição (Grünbaum, 1985). Mesmo assim, essa evidência contrária não implica na falsidade de todo o edifício teórico da Psicanálise, pois põe em cheque apenas uma parte que, mesmo fundamental, é relativamente independente de outras regiões da teoria.

Mesmo contrária à concepção de Popper, a posição de Grünbaum é ainda mais mordaz. Ao mesmo tempo em que reconhece o potencial científico da Psicanálise, pelo menos com relação ao critério de demarcação de Popper, Grünbaum aponta para a irrisória base de evidências da eficácia do tratamento psicanalítico e das próprias teorias metapsicológicas. Isso porque a principal base de evidência provém do relato de casos e excertos clínicos, cuja validade e veracidade depende não só da honestidade do analista ao relatar as informações de forma fidedigna e isenta de seleções parciais, mas também da ausência de contaminação das associações do paciente (Grünbaum, 1985).

A própria teoria psicanalítica, contudo, admite a vital importância da sugestão no processo de análise, operacionalizando parte dessa sugestão sob o conceito de transferência. Como avaliar o quanto o foco do paciente num conjunto de memórias não é nada mais do que uma tentativa de agradar a clara predileção do analista por determinados tipos de narrativas? O quanto o uso de breves intervenções ou silêncios serve de sinal para o paciente enunciar mais determinados conteúdos e não outros? A justificativa dada por Freud, batizada por Grünbaum de *tally argument*, supõe que a sugestão pode até influenciar intelectualmente o paciente, mas somente uma interpretação que de fato reflita o conteúdo recalcado irá surtir efeito sobre seus sintomas. A justificativa é sua própria derrocada: se as melhoras exibidas pelo paciente e sua confirmação do acerto da interpretação podem provir da sugestão do analista, utilizá-las como evidência confirmatória do bom andamento do tratamento e das conjecturas psicanalíticas sobre o funcionamento da psique é cair numa circularidade insolúvel.

A desconstrução feroz do valor dos dados clínicos empreendida por Grünbaum fez surgir réplicas reiterando o valor dos estudos de caso a partir da clínica, com argumentos variando de justificativas filosóficas para seu fundamento (Mezan, 2006), ou propondo novos critérios de validade para dados clínicos (Edelson, 1985), até procedimentos intersubjetivos de análise baseados em procedimentos bem delimitados para rastrear evidência de temáticas pertinentes às sessões de um estudo de caso (Meehl, 1983). Por outro lado, a crítica de Grünbaum fez crescer as pesquisas psicanalíticas com dados provenientes de fora das sessões clínicas. Em defesa da clínica ou buscando novos dados fora dela, há pouca evidência de que os estudos psicanalíticos empreendidos nos contextos universitários ajudam a corroborar ou mesmo a pôr em cheque aspectos relevan-

tes da teoria psicanalítica, colocando a Psicanálise na condição de um programa de pesquisa degenerado (Lakatos, Worrall, & Currie, 1980). Parte dos estudos empíricos internacionais está voltada para avaliar os efeitos clínicos do tratamento, mas há alguns desenvolvimentos, em menor escala, que buscam inovações para melhorar a operacionalização e, conseqüentemente, a testagem de elementos da própria teoria psicanalítica (Leuzinger-Bohleber & Kächele, 2015). No contexto brasileiro, a maioria dos trabalhos desenvolvidos são de caráter teórico, seguido de perto por pesquisas interessadas na interpretação de elementos culturais e sociais – ambas ancoradas na tradição hermenêutica que busca compreender um dado fenômeno a partir de uma chave interpretativa (Blatt, Corveleyn, & Luyten, 2006). Na contracorrente dessa tendência, há raros estudos empíricos focados principalmente na investigação do processo terapêutico, como a avaliação da dinâmica entre terapeuta e paciente (Serralta, 2016) por meio de instrumentos padronizados, como o *Psychotherapy Process Q-set* (Serralta, Nunes, & Eizirik, 2007) e avaliação de indicadores sobre abandono do tratamento (S. I. Jung et al., 2014).

Partindo desse panorama, fica evidente a necessidade constante de criar e avaliar procedimentos empíricos bem definidos para tornar mais robustas as inferências realizadas a partir de pesquisas em Psicanálise. Se os psicanalistas muitas vezes se eximem de pôr em questão a evidência utilizada para sustentar a Psicanálise como prática e teoria, seus detratores o fazem com um fervor considerável e, frequentemente, com uma ignorância em igual tamanho. Precisamos, portanto, fazer retornar para a própria Psicanálise a questão sobre o que constitui evidência em seu campo de estudo, respondendo à altura das críticas de Popper e Grünbaum, e avançando a partir dos esforços de Dahl, Meehl, Thomä e Kächele – já um pouco esquecidos ou, no contexto brasileiro, praticamente desconhecidos.

## Modelos analógicos e digitais

Frente às críticas expostas acima, vemo-nos confrontados com três desafios distintos. Em primeiro lugar, retomando a crítica de Popper, é necessário deduzir das teorias psicanalíticas enunciados bem definidos e relativamente arriscados. Dificuldades nessa dedução podem implicar em reformulações das teorias para permitir sua operacionalização. Em segundo lugar, essas predições só poderão ser postas à prova se pudermos decidir por uma fonte de evidências confiável dos processos psíquicos. A crítica de Grünbaum põe em

cheque o uso de dados provenientes da clínica, o que pode sugerir a necessidade de buscar evidências em outros contextos. De nossa parte, não desconsideramos os limites da evidência clínica indicados por Grünbaum – sua contaminação por sugestões do analista, sua profusão e, portanto, difícil manejo inferencial – mas nos mantemos alinhados à posição de Meehl (1983): a despeito da importância de outras fontes de evidência, as sessões clínicas são o melhor contexto para testar hipóteses psicanalíticas. Essa postura, contudo, exige que o valor probatório da clínica seja demonstrado, e não pode ser justificado simplesmente como postulado. O que nos leva ao terceiro desafio, imediatamente derivado do segundo: se escolhemos manter a clínica como referência preferencial para avaliar as hipóteses deduzidas das teorias psicanalíticas, é necessário desenvolver procedimentos que reduzam a discricionariedade do analista na seleção do material, tornem evidentes possíveis efeitos de sugestão, e se baseiem numa representação mais integral dos eventos clínicos, permitindo trabalhar as associações o mais próximo de sua totalidade e evitando o recurso ao *tally argument* como critério para validação das hipóteses clínicas.

Esses três desafios são atravessados por uma característica da teoria psicanalítica que os tornam particularmente problemáticos. A dificuldade em deduzir hipóteses testáveis, avaliar a fidedignidade da experiência clínica e utilizá-la como evidência podem ser atribuídas à utilização, por Freud e a grande maioria dos psicanalistas, de modelos analógicos para compreensão de fenômenos metapsicológicos. Entendemos aqui modelos analógicos no sentido dado por Bunge (1974) aos suportes utilizados para descrever de maneira intuitiva os objetos modelos abordados por uma teoria específica. Esses suportes, como figuras, animações, esquemas e analogias, têm função psicológica importante para a transmissibilidade e compreensão dos objetos modelos, mas não lhe podem servir de substitutos porque sua natureza figurativa impede a descrição sistemática por meio de modelos teóricos. Como evidencia Mahony (1990) na análise do estilo da escrita freudiana, Freud emprega constantemente recursos figurativos para estabelecer analogias entre fenômenos psíquicos latentes e sistemas físicos conhecidos.

Lançando mão da metáfora para traduzir num sistema conhecido os processos subjetivos inacessíveis, Freud constrói modelos analógicos que suprem diversas funções. Por vezes, esses modelos servem à função pedagógica de explicar um fenômeno dinâmico, como o modelo do “ouvinte hostil” proferido nas *Cinco lições de*



*Psicanálise* para representar o processo de recalque e retorno do recalcado. Outras vezes, o modelo analógico se presta como heurística para ensaiar uma explicação sobre o funcionamento psíquico, como o modelo da lamela unicelular, empregado no *Além do princípio de prazer* para compreender a origem da pulsão de morte. Por fim, muitos modelos são tomados como estruturais, quando são utilizados como uma descrição ontológica da substância e processos psíquicos em si, como a conjugação de traços mnêmicos registrados em duas instâncias qualitativamente diferentes e investidos de *quanta* de energia que explicaria o funcionamento do recalque.

Por que esses modelos podem servir de barreira na pesquisa psicanalítica? Presos a esses modelos, não temos outra opção senão utilizarmos dessas categorias para interpretar, no sentido hermenêutico, quaisquer dados que desejemos analisar – sejam relatos clínicos, fenômenos sociais ou a própria teoria analítica. Por não comportarem uma descrição bem definida de suas propriedades, os modelos analógicos não se prestam a uma operacionalização direta. Sua utilização sempre é aplicada no sentido tradicionalmente apontado para as Ciências do Espírito: compreender retroativamente aplicando uma chave interpretativa que permita reduzir os dados a categorias conhecidas (Blatt et al., 2006; Mezan, 2007). Sem uma operacionalização direta, é difícil deduzir hipóteses precisas e arriscadas e, dessa forma, pôr o modelo à prova e, possivelmente, refutá-lo. E como cada chave interpretativa é desenvolvida com interesses específicos, ao sabor da subjetividade do pesquisador, agregar diferentes estudos para buscar conclusões mais abrangentes é quase impossível.

Trata-se, em última instância, de um problema epistemológico. Primeiramente, o objeto modelo e sua especificação não são definidos com clareza. Até que ponto é possível levar uma representação analógica? O uso da analogia se propõe aproximativa ou pretende representar de maneira mais direta os fenômenos em si? As analogias e metáforas propostas por Freud transitam de meros recursos heurísticos a ensaios de formalização – basta comparar a diferença que separa o modelo da lamela da descrição do funcionamento do recalque e do inconsciente em seus escritos homônimos. Mas mesmo quando há a produção de objetos modelos do funcionamento psíquico – como as noções de representações de coisa e de palavra, energias livre e ligada – o modelo teórico que as descreve ainda é excessivamente vago para garantir a produção de um sistema hipotético-dedutivo.

O que propomos em seu lugar é a transformação de alguns desses modelos analógicos – ou aspectos essenciais desses modelos – em *modelos digitais*. A noção de modelo digital aqui é a tomada de empréstimo de Bunge (1973), a partir da maneira como ele descreve as teorias específicas de objetos modelos. Para Bunge, as ciências não lidam diretamente com os objetos e fenômenos reais sobre os quais se debruçam. Apesar de manter uma relação com um determinado recorte da realidade, as ciências operam com representações abstratas e idealizadas desses recortes. O processo de abstração retira da realidade as características essenciais para a constituição do modelo, descartando aspectos considerados acessórios. Como o modelo é aproximativo, o que é considerado essencial e acessório num momento pode ser revisado posteriormente, sob a luz de novas evidências. As abstrações são então sintetizadas numa representação ideal que forma o objeto modelo.

Até aqui, um objeto modelo pode ser tanto digital quanto analógico. Afinal de contas, um modelo analógico também funciona abstraindo características relevantes e formulando idealizações apoiadas em outro fenômeno similar e mais conhecido. Mas um objeto modelo só pode ser efetivamente operacionalizado – isto é, utilizado para fazer inferências sobre a realidade – na medida em que eles são descritos por meio de uma teoria específica, com auxílio de uma teoria geral já elaborada (Machado & Cruz, 2011). A teorização explícita de um objeto modelo se constitui, em última instância, como um modelo matemático da realidade. O recurso à matemática não se deve apenas a um apelo ao formalismo, mas à capacidade que um objeto modelo, matematicamente descrito, tem de gerar deduções decorrentes de teorias matemáticas gerais já estabelecidas.

É aqui que distinguimos os modelos analógicos dos modelos digitais. Um modelo analógico teoriza sobre um objeto modelo inserindo-o num quadro de referências familiar por analogia. Seu poder explicativo é qualitativo e dado pela alcance de sobreposição entre o modelo representado e seu análogo. Um modelo digital, por outro lado, não se baseia simplesmente na similaridade analógica entre diferentes sistemas, mas faz uso de conceitos matemáticos para representar o objeto modelo.

A utilização da matemática traz uma série de vantagens: podemos fazer uso de objetos já bem definidos e formalizados em diferentes teorias matemáticas, o que facilita realizar operações a partir deles; a representação formal diminui a ambiguidade da teoria específica, tornando-a mais facilmente comunicável; sendo imedia-

tamente operacionalizáveis, os modelos digitais podem produzir resultados com maior coerência entre teoria e dados empíricos, além de tornarem mais comparáveis os resultados de diferentes pesquisas. Há limites para sua utilização, é claro: somos obrigados a utilizar os objetos matemáticos já conhecidos, o que nos exige adequar nossos fenômenos de interesse a esses objetos – a não ser que estejamos dispostos a criar novos objetos formais; um modelo digital precisa necessariamente delimitar seu escopo para ser funcional e utilizável, reduzindo sua abrangência quando comparado a um modelo analógico mais geral.

Acreditamos, porém, que os benefícios trazidos pelo emprego de modelos digitais os tornam preferíveis ao emprego exclusivo de modelos analógicos – o que não implica em abandonar o uso de modelos analógicos por completo.

## Por um modelo digital das associações livres

Mas que tipo de modelo digital poderia ser utilizado para representar qualquer aspecto da Psicanálise? É difícil argumentar que a Psicanálise constitui uma única teoria geral. Pelo contrário, conjecturas diversas tentam dar inteligibilidade a diferentes fenômenos psíquicos, e essas conjecturas nem sempre se encaixam numa teoria mais ampla. Esse é um argumento dado por Grünbaum (1985) para justificar que, mesmo que a Psicanálise produza previsões arriscadas e encontre um contexto de justificação, dificilmente seu edifício teórico seria posto à prova por completo.

Precisamos, portanto, escolher algum aspecto relevante dentre todos os possíveis. Se temos por objetivo demonstrar que evidências provenientes da clínica podem servir para corroborar hipóteses psicanalíticas, é necessário se focar nos principais procedimentos práticos utilizados pelos psicanalistas. Ora, a prática da psicanálise se distingue de outras psicoterapias principalmente pelo uso da *regra fundamental*, a associação livre (para o analisando), e seu correlato, a atenção livremente flutuante (para o analista). Toda e qualquer inferência feita sobre os processos psíquicos a partir da experiência clínica deve ser remontado, com maior ou menor grau de abstração, às cadeias associativas pronunciadas pelo analisando durante suas sessões. Da parte do analista, a conjugação da associação livre à escuta flutuante tem por finalidade permitir a elaboração de

*interpretações*. Como, porém, a escuta das associações de um analisando permite interpretar os derivados do inconsciente? Como será visto no primeiro capítulo, interpretar em Psicanálise não é atribuir sentido, como o hermenêuta que reduz o mundo às suas categorias interpretativas, mas reconstituir o lugar dos derivados do inconsciente na experiência subjetiva do sujeito por meio de suas cadeias associativas. Interpretar, em última instância, é descobrir os pontos de convergência nas associações livres de forma a reestabelecer a continuidade da experiência consciente. Por isso, mais do que uma ferramenta de intervenção clínica, as interpretações são hipóteses sobre a organização do psiquismo de um sujeito, permitindo relacioná-lo com o psiquismo de modo geral.

Interessa-nos, portanto, construir conjuntamente uma representação abstrata das associações livres, tal qual elas surgem no contexto clínico, e uma representação esquemática dos elementos psíquicos que determinam o surgimento das associações a partir de complexos inconscientes latentes. Nosso modelo teórico consistirá em descrever exatamente como esses complexos latentes produzem as associações escutadas, tornando possível construir procedimentos de inferência no sentido inverso, tal como a interpretação dos sonhos permite inferir os pensamentos oníricos a partir do conteúdo manifesto, mediante a reconstrução da elaboração onírica.

O processo pelo qual é possível inferir qualquer coisa sobre os processos inconscientes foi justamente o que motivou o rompimento entre Freud e Fliess, como exposto acima. Sua centralidade é reconhecida por Freud: se não podemos sustentar que esse procedimento é dotado de alguma objetividade, todo o empreendimento analítico está de saída fadado ao fracasso, pois estaria limitado a descrever apenas a subjetividade do pesquisador. Essa preocupação também levou Meehl (1983) a propor uma série de alternativas para melhorar o estatuto da evidência proveniente das sessões clínicas. Concordamos com ele que as sessões continuam sendo, ainda, a melhor fonte de evidência para a pesquisa em Psicanálise, apesar da maneira muitas vezes questionável com a qual elas são tratadas nas pesquisas em Psicanálise e de suas óbvias limitações com relações a questões teóricas mais gerais. Insistindo que os dados brutos provenientes da clínica se reduzem, exclusivamente, a enunciados, gestos e posturas, Meehl sugere cinco procedimentos que podem auxiliar na redução da subjetividade nas pesquisas que fazem uso de dados clínicos:

1. Aplicação de estatísticas simples a evidências diretas de pro-

- cessos psicanalíticos de superfície retirados da prática clínica;
2. Utilização de técnicas estatísticas mais complexas, como análise fatorial e de *clusters*;
  3. Apropriação de novos formalismos matemáticos, adequados ao fenômeno em questão – a sessão clínica;
  4. Desenvolvimento de programas de computador para análise de conteúdo avançada, como o *General Inquirer*;
  5. Buscar por temas evidentes a partir de blocos de tópicos por meio de um procedimento inovador que ele denomina de *Topic Block Theme Tracing* (rastreamento temático por blocos de tópicos).

Apesar das diferenças metodológicas contidas nas propostas, todas elas têm como denominador comum o estabelecimento de procedimentos que permitam relacionar um aspecto empírico passível de ser registrado na clínica – os enunciados que emergem das associações livres do paciente – com a construção que o psicanalista efetua, nem sempre de modo consciente, a respeito da ligação entre representações inconscientes no psiquismo do analisando.

Inspirados pelas sugestões de Meehl, proporemos, então, um modelo digital para representar o processo gerador das associações livres na clínica analítica, de forma a permitir chegar a seus determinantes nos complexos inconscientes, de maneira semelhante à qual o analista chega a suas interpretações. Como poderemos criar um modelo para esse processo? Felizmente, não somos obrigados a partir do zero. Há, na área de Inteligência Artificial, modelos probabilísticos que buscam inferir a estrutura latente de núcleos semânticos a partir de dados textuais. Esses modelos são chamados de *modelos de tópicos* e serão abordados em profundidade no segundo capítulo. Dessa maneira, utilizaremos objetos provenientes da teoria da probabilidade para formalizar o aspecto digital de nosso modelo.

Esperamos, assim, conjugar as cinco sugestões feitas por Meehl (1983, p. 380) para diminuir a subjetividade do pesquisador ao utilizar dados provenientes das sessões clínicas: nosso objeto modelo será construído como uma abstração de fenômenos de superfície, ou seja, eventos observáveis e registráveis que acontecem durante uma sessão, de forma que nosso formalismo não exija grande construções sobre os fenômenos representados; utilizaremos uma técnica estatística complexa, os modelos de tópicos, cujo objetivo é a redução

da dimensionalidade dos dados, e por isso próxima da análise fatorial; aplicaremos um formalismo matemático que, apesar de não ser completamente novo, sintetiza contribuições inovadoras da teoria da probabilidade, da teoria dos grafos e das Ciências da Computação; utilizaremos um programa de computador para auxiliar a tarefa investigativa, única maneira praticável de se analisar um grande volume de dados e tornar possível a reprodução das análises conduzidas; e, por fim, desenvolveremos uma alternativa à sua proposta original no artigo de 1983, o *rastreamento de temas por blocos de tópicos*, a partir de um procedimento não supervisionado de aprendizagem estatística, que permita entrever a trama de tópicos abordados pelo paciente ao longo de sua análise.

## Escopo da tese

Nossa tese, portanto, é esta: *o processo de associação livre pode ser representado como um modelo de tópico*. Isso significa que, partindo dos modelos analógicos de Freud e dos modelos probabilísticos de tópicos, desenvolvemos um modelo digital representativo do processo de associação livre como determinado por complexos inconscientes. Analisamos um caso clínico para ilustrar o potencial da utilização dos modelos de tópicos para a redução de dados e a criação de sumários significativos a partir de transcrições de sessões clínicas. Por fim, o modelo é posto à prova no sentido de seu poder preditivo para novas sessões e para prever o resultado de testes de associação de palavras.

Com isso, assumimos uma postura específica para responder à questão do lugar da pesquisa em Psicanálise na universidade. A pesquisa psicanalítica pode e deve se basear em evidências obtidas a partir da experiência clínica. Essas evidências, contudo, não são apenas os dados clínicos brutos ou seleções feitas ao bel-prazer do pesquisador, mas relações inferidas a partir da utilização de nosso modelo digital. Esperamos, dessa forma, não só indicar uma nova proposta metodológica para a pesquisa em Psicanálise, mas também insistir na possibilidade da utilização de modelos digitais para formalizar aspectos da teoria analítica. Por fim, esperamos demonstrar a versatilidade de nossa proposta em comunicar evidências provenientes da experiência clínica, facilitando o diálogo entre pesquisadores e permitindo melhorar a qualidade da evidência nas publicações da área.

Iniciamos o percurso da tese retomando os fundamentos me-

tapsicológicos do processo de associação livre na literatura psicanalítica. Apesar de sua importância para a técnica e para a pesquisa psicanalítica, a metapsicologia das associações livres ainda é deficitária, e por isso rastreamos as justificativas dadas por Freud para utilizá-la em suas investigações sobre o inconsciente.

A seção seguinte apresenta em profundidade o que são modelos de tópicos. Como as bases matemáticas utilizadas por esses modelos são pouco conhecidas dentro da Psicologia, o segundo capítulo aborda também aspectos básicos sobre a inferência bayesiana, a construção de modelos gráficos probabilísticos e o modelo de tópico arquetípico: a *Alocação de Dirichlet Latente* (*Latent Dirichlet Allocation*, LDA). Encerramos a seção apresentando estudos da área da Psicologia que já fazem uso de modelos de tópicos, seja como ferramenta exploratória, seja como modelo formal para fenômenos psicológicos.

Em seguida, no capítulo três, tendo já esclarecido o modelo analógico do qual partimos, retomamos as definições formais dos modelos de tópicos para propor um modelo digital das associações livres. Discutimos o problema de como implementar um algoritmo capaz de fazer uso das transcrições das sessões clínicas, e indicamos os atalhos que precisaram ser tomados para tornar a implementação possível. Abordamos também como foi desenvolvido o estudo de caso que serve de exemplar para a aplicação de nosso modelo.

Discutimos como os dados das transcrições clínicas foram tratados no capítulo quatro. Também abordamos os detalhes sobre a escolha do número de tópicos e do modelo com melhor ajuste. Apresentamos os critérios objetivos utilizados para comparar diferentes ajustes entre si, analisando o impacto dos algoritmos utilizados sobre o modelo final. Avaliamos as evidências de validade do modelo por meio de duas estratégias de validação. Primeiro, com base na avaliação do ajuste do modelo aos pressupostos por meio da técnica de checagem preditiva posterior. Em seguida, apresentamos o desempenho do modelo escolhido para a predição das palavras mais prováveis de serem associadas em um teste de associação de palavras.

O último capítulo apresenta as principais inferências que podemos fazer a partir do modelo de tópicos utilizado para modelar as associações livres do caso estudado. Focaremos nos conteúdos dos tópicos descobertos, utilizando excertos das sessões para compreender melhor seu escopo. Avaliamos também a dinâmica temporal dos tópicos, procurando identificar tendências de redução ou aumento

de determinadas temáticas.



# Capítulo 1

## Associações livres

### 1.1 Nossos antecedentes

É quase inexistente, em toda a literatura da Psicanálise, a proposição de modelos matemáticos como complemento ou formalização dos modelos analógicos propostos por Freud e por outros psicanalistas. O que podemos encontrar mais próximo disso é a exigência da avaliação empírica das teorias psicanalíticas, como na relativamente desconhecida obra do psicanalista estado-unidense Hartvig Dahl (Dahl et al., 1988) e o trabalho conjunto dos alemães Helmut Thomä e Horst Kächele, membros do *Grupo de Estudos sobre Pesquisa de Processos Psicanalíticos de Ulm* (Kächele et al., 2011).

Hartvig Dahl foi um psicanalista norte-americano que propôs, iniciando na década de 70, o desenvolvimento de procedimentos de investigação empírica para tornar mais rigorosa a fundamentação da técnica e da teoria psicanalítica. Ele também foi pioneiro na aplicação de procedimentos computadorizados para a análise de dados provenientes da clínica, como transcrições de sessões, chegando a apresentar análises de casos cujas sessões foram transcritas integralmente. Um dos principais casos foi a psicanálise de Mrs. C, uma jovem obsessiva de 28 anos de idade que teve toda as suas 363 sessões gravadas em áudio. As transcrições dessas sessões serviram de base para várias proposições teórico-práticas de Dahl.

Dentre essas contribuições teóricas está a descrição da ocorrência de *quadros* (*Frames*, no original, Dahl et al. 1988, posteriormente tornado um acrônimo de *Fundamental, Repetitive And Maladaptive Emotional Structures*) como estruturas de conhecimento estereotipadas e associadas a emoções que se repetem nas associações livres do analisando. Esses diferentes quadros servem de base prototípica para a instanciação de diferentes associações. Como os quadros são instanciados pelas associações, é possível retirar da transcrição de sessões clínicas evidências de sua existência e utilizá-

-los para descrever os conflitos que emergem no processo psicanalítico, ainda que a proposição de que diferentes instâncias pertencem ao mesmo protótipo repetitivo permaneça, em boa parte, sob o jugo do pesquisador ou de juízes especialistas. Junto com os avanços teóricos, Dahl e sua esposa, Virginia Teller, propuseram também a utilização de técnicas de análise de conteúdo computadorizadas, baseadas na utilização de um dicionário de termos relevantes e paudadas principalmente na frequência de ocorrência e coocorrência dos termos julgados importantes *a priori*, com tratamentos posteriores utilizando análise fatorial e de agrupamentos. Teller, em particular, por trabalhar diretamente na área de computação com técnicas de análise de linguagem natural, argumentou em favor da importância dos estudos da área de Inteligência Artificial para a construção de modelos informatizados do funcionamento do psiquismo – projeto que, infelizmente, não avançou para além de esboços e ensaios (Dahl et al., 1988).

As limitações na tecnologia informática da época, contudo, e a divulgação reduzida dos resultados de suas pesquisas acabou por tornar suas contribuições pouco reconhecidas, mesmo dentro da Psicanálise. De fato, o pequeno impacto de seus trabalhos pode também ser atribuído à resistência, entre alguns pesquisadores da Psicanálise, à formalização e à utilização de recursos empíricos mais rigorosos, a ponto de Kächele denominá-lo de “cavaleiro solitário” em sua elegia em função de sua postura *sui generis* a respeito dos procedimentos de pesquisa em Psicanálise (Kächele & Hölzer, 2007).

A base dessa resistência pode ser encontrada tanto no desconhecimento das possibilidades de investigação empírica quanto nas transformações de certas recomendações feitas sobre a técnica da Psicanálise em uma ortodoxia inquestionável. Essa ortodoxia é discutida e refutada, por exemplo, no caso do uso de instrumentos para gravação de áudio durante as sessões de análise pela equipe de pesquisa de Thomä e Kächele (Dahl et al., 1988). O *Grupo de Estudos Sobre Pesquisa de Processos Psicanalíticos de Ulm* foi, de fato, um dos poucos interlocutores de Hartvig Dahl e um grupo expoente na pesquisa empírica sobre o processo psicanalítico. Para o Grupo de Ulm, há diferenças fundamentais entre a pesquisa sobre o *processo* da análise e sobre os *desfechos* de um tratamento. Uma pesquisa sobre possíveis desfechos de um tratamento pode seguir os moldes dos ensaios clínicos já tradicionais na medicina: estipula-se um desfecho principal, randomiza-se o tratamento e comparam-se as diferenças entre grupos experimentais e controle. O estudo do processo, foco

do interesse do Grupo de Ulm, necessita de outro tipo de evidência, capaz de permitir inferências causais entre os eventos vivenciados na clínica e os desfechos observados. Essa evidência é buscada na transcrição direta das sessões clínicas – uma exigência em comum com a proposta de Dahl – e avaliada por meio de técnicas de exploração de dados textuais, como contagem de termos e grupos de termos, operacionalizados, por exemplo, no *Ulm Textbank*, e muitas vezes modelados utilizando técnicas estatísticas tradicionais, como séries temporais (Dahl et al., 1988).

A obra de Thomä, Kächele e seus colaboradores do Grupo de Ulm alcançaram uma maior notoriedade internacional, inclusive com um volume de seus trabalhos publicado em português (Thomä & Kächele, 1992), e culminando na realização de um evento conjunto dos membros do grupo com Dahl e seus colaboradores (Dahl et al., 1988). Mesmo assim, os resultados de suas pesquisas permanecem pouco organizados e reconhecidos pela comunidade psicanalítica, a ponto de uma recente discussão sobre a pertinência de estudos de casos psicanalíticos no *British Journal of Psychiatry* (Wolpert & Fonagy, 2009) sequer mencionar seus nomes na defesa dos estudos de caso. O desaparecimento e esquecimento das pesquisas empíricas mais rigorosas em Psicanálise parece ser a regra, pois, mesmo ecoando as preocupações de Dahl, Thomä e Kächele, e criticando a posição pouco científica de muitos psicanalistas, as alternativas propostas por Meehl (1983) para a pesquisa com o material clínico, como apresentamos, a título de inspiração, na Introdução, não foram levadas a cabo e também são pouco conhecidas.

O ponto em comum de nossos autores de referência é a exigência da fundamentação da pesquisa em Psicanálise em dados confiáveis provenientes da prática clínica. Dahl, Thomä e Kächele são inflexíveis quanto à necessidade desse material ter origem na gravação e transcrição direta das *associações livres* dos analisandos. A matéria-prima da pesquisa em Psicanálise deve ser, preferencialmente, o mesmo material com o qual o psicanalista trabalha quando conduz uma análise – e de onde Freud elaborou seus constructos teóricos. Os autores também estão interessados em investigar os *processos* que se desenrolam ao longo de uma análise, e não somente a avaliação dos desfechos – ainda que Kächele também tenha participado ativamente de pesquisas voltadas para a avaliação da eficácia terapêutica da Psicanálise (Leuzinger-Bohleber & Kächele, 2015). Um outro ponto em comum é a rejeição das leituras hermenêuticas da Psicanálise. Seguindo a definição de Freud (1984k) de que a Psicaná-

lise é uma *Naturwissenschaft* (como discutido também em Mezan, 2007), esses autores concordam que o trabalho de pesquisa puramente interpretativo não contribui para uma boa fundamentação científica da Psicanálise, e que esse fundamento só pode ser alcançado por meio da construção de hipóteses e sua avaliação por meio de pesquisas empíricas. Nenhum deles, todavia, chega a propor um modelo propriamente *matemático*: as diferentes proposições teóricas são ainda excessivamente esquemáticas e pouco descritas em termos de teorias específicas. Em suma, a matemática ainda cumpre o papel denunciado por Bunge (1974) de uma mera heurística para resumo e análise do grande volume de dados empíricos, uma ponte frágil entre os modelos analógicos tradicionais da Psicanálise e os modelos digitais comuns nas ciências naturais.

## 1.2 Origem do método de associação livre

Alertados sobre os riscos da pesquisa com dados clínicos, conforme Grünbaum, e inspirados nas propostas de Meehl e no programa de pesquisa iniciado por Dahl, Thomä e Kächele, propomos construir um modelo digital que permita firmar um compromisso entre alguns aspectos essenciais da teoria e da prática psicanalítica, mas que seja acessível o suficiente para ser submetido a uma investigação empírica. Esse modelo deve abordar aspectos cruciais de um fenômeno que serve de ponte entre a metapsicologia – o edifício conceitual que Freud elaborou a partir de sua experiência clínica para organizar suas descobertas – e a prática clínica tradicional concebida a partir da *regra fundamental da Psicanálise* (Freud, 1984d): o processo de *livre associação*.

O processo de livre associação (*freier Einfall*) já se apresentava de maneira embrionária no método catártico. O grande avanço desse método sobre o uso já reconhecido da sugestão hipnótica era, justamente, permitir que o paciente hipnotizado recordasse de lembranças ligadas a eventos traumáticos, em vez de simplesmente obedecer as sugestões do hipnotizador (Breuer & Freud, 1984). Freud, porém, vê problemas na dependência do tratamento catártico com relação a hipnose, e a abandona. O abandono da hipnose foi motivado por problemas técnicos e éticos (Aguiar, 2016): por questões técnicas, na medida em que alguns pacientes eram pouco sugestionáveis, difíceis de serem induzidos ao estado hipnótico; porque os resultados do método catártico eram pouco estáveis, com melhoras notáveis logo cedendo lugar a recaídas; e por questões éticas, na

medida em que o processo sugestivo da hipnose implicava na dominação do hipnotizado pelo hipnotizador e na desconsideração da seriedade da patologia, na medida em que o sofrimento do paciente era reduzido a sugestões de que não lhe acometia nenhuma doença de verdade. As experiências de sugestão pós-hipnótica empreendidas por Bernheim deram a Freud segurança de que as memórias facilmente resgatadas durante o estado hipnótico também poderiam ser resgatadas em vigília, desde que o terapeuta insistisse em sua recuperação. Surge aí um procedimento transicional entre o método catártico e o método psicanalítico propriamente dito: a rememoração por atenção concentrada (Freud, 1984i).

A associação livre surge, finalmente, quando a utilização de um estado alterado de consciência e a busca intencional de elementos patógenos sob pressão do analista cede lugar à expressão livre do paciente (Freud, 1984i). Essa passagem não se dá apenas por uma escolha metodológica de Freud, mas é inspirada na sugestão feita por uma de suas pacientes, Frau Emmy von N, que insiste para que ele a deixe falar livremente e abandone a insistência de recordar episódios específicos (Laplanche & Pontalis, 1998). Freud também encontra antecessores no uso de associações livres no médico-místico britânico Garth Wilkinson e no poeta alemão Friederich Schiller, mas reconhece que sua proposição por esses autores era com fins da produção literária, e que não o influenciaram diretamente. A influência direta veio de outra fonte: os ensaios do escritor alemão Ludwig Börne. Freud afirma ter lido algumas de suas obras quando tinha 14 anos, e essas leituras o marcaram profundamente. Apesar de não recordar do pequeno ensaio *A arte de tornar-se um escritor original em três dias*, onde Börne sugere o método de associação livre como ferramenta para a escrita, Freud (1984j) admite lembrar-se de vários de seus contos e conselhos.

Na prática clínica, o método de livre associação consiste numa proposição simples: o analista pede para que o paciente abandone qualquer intenção consciente e simplesmente fale o que lhe ocorre em seus pensamentos, independente de qualquer crítica que possa lhe surgir (Freud, 1984g, 1984i). Esse procedimento se torna particularmente evidente a partir da descrição da técnica de análise dos sonhos. Na interpretação dos sonhos, elementos do conteúdo manifesto servem de ponto de partida para a formação de cadeias associativas que permitem desvendar os pensamentos oníricos latentes (Freud, 2001).

A utilidade do método de associação livre para a Psicaná-

lise não se deve unicamente em permitir ampliar o tratamento para pacientes refratários à hipnose – limitação que Freud identificava desde seu estágio com Charcot e de sua experiência com Bernheim, em Nancy – mas também porque as associações permitiram lançar novas possibilidades ao tratamento psicanalítico, como a dissolução de resistências, ao mesmo tempo em que se mostravam consonante com a compreensão de Freud sobre o funcionamento do psiquismo. Apesar de sua importância, é notável que uma metapsicologia das associações livres não ocupe o mesmo lugar de mérito como a descrição do recalque, da transferência ou da angústia. Uma breve busca na edição *Standard* brasileira das obras de Freud<sup>1</sup> revela que a locução “associação livre” ou “associações livres” ocorrem apenas 137 vezes ao longo de toda a obra, enquanto “transferência”, a título de comparação, aparece cerca de 648 vezes.

Faremos, portanto, uma breve retomada dos fundamentos metapsicológicos para o uso das associações livres na clínica psicanalítica. Com auxílio dessa fundamentação, poderemos passar ao ponto central desta tese: construir um modelo formal do processo de livre associação baseado nos modelos de tópicos. A partir desse modelo, por fim, poderemos abordar o trabalho empírico desenvolvido a partir de um estudo de caso.

### 1.3 Modelos analógicos apresentados por Freud

A importância do método de associação livre para a técnica psicanalítica é reiterada em vários lugares na obra de Freud, e particularmente nos escritos sobre a técnica, nos quais é denominada de “regra fundamental da Psicanálise” ou “regra de ouro da Psicanálise” (Freud, 1984d). Sua importância na prática psicanalítica tradicional é um ponto pacífico e não será desenvolvido em maiores detalhes. No lugar de discutir a pertinência das associações livres

---

<sup>1</sup>Convém discutir brevemente a escolha da edição *Standard* brasileira da Imago como referência para os escritos de Freud. A despeito de suas conhecidas limitações em termos de problemas editoriais e, sobretudo, de tradução, a edição *Standard* da Imago é, de longe, a mais acessível e mais lida versão da tradução dos escritos de Freud no Brasil. Com isso, esperamos facilitar para os leitores encontrar as referências citadas. Acresce ainda o fato de que nosso objeto de interesse – as associações livres – não apresenta questões de tradução particularmente perniciosas, podendo prescindir das filigranas semânticas que orientam a escolha de diferentes tradutores.

como principal meio para exploração do psiquismo, retomaremos a fundamentação teórica para seu uso como procedimento central do tratamento psicanalítico. Retomamos os escritos freudianos, portanto, com a seguinte pergunta: quais as garantias ou justificativas que Freud dá de que as associações livres permitem, de fato, lançar alguma luz sobre os processos psíquicos e, em especial, os processos psíquicos inconscientes? Infelizmente, ao longo da obra de Freud, os fundamentos metapsicológicos das associações livres são esparsos. Por isso, abordaremos sua principal elaboração metapsicológica, apresentada no capítulo VII de *A interpretação dos sonhos*.

Freud resume sua principal descoberta em *A interpretação dos sonhos* como o reconhecimento de que o sonho é um “processo dotado de sentido e passível de ser inserido na cadeia de experiências psíquicas do sonhador” (Freud, 2001, p. 493). A tese de que o sonho é um fenômeno psíquico significativo é o alvo central de toda a obra, mas o complemento dado por Freud aqui nos interessa bastante. Reconhecer o sentido do sonho não é descobrir uma chave hermenêutica para desvendá-lo, mas implica, como já referido, em inseri-lo na cadeia de experiências psíquicas do sujeito. Ora, uma de suas metáforas para designar o material inconsciente que consegue galgar seu caminho até a consciência é *lacuna*, ou seja, lapsos da experiência consciente do sujeito que são atravessados por processos cuja origem lhe escapa (Freud, 1984e). Interpretar um sonho – ou qualquer outro derivado do inconsciente – implica em conseguir reconstruir os elos perdidos entre o material manifesto, lacunar, e a continuidade da experiência psíquica. Laplanche (2014) insiste, inclusive, que interpretar em Psicanálise não é encontrar uma chave oculta que permite verter um código em sua mensagem subjacente, mas reconstruir os fios discursivos que *sobredeterminam* a experiência psíquica manifesta.

A noção de que é possível reconstituir o lugar dos derivados do inconsciente com relação ao restante da vida psíquica é tributária do forte valor dado por Freud ao determinismo dos fenômenos psíquicos. Partindo do determinismo como axioma, não é possível conceber que qualquer fenômeno psíquico, mesmo desprovido de importância imediata, poderia ocorrer senão em função de uma cadeia de eventos que o causaram. Freud admite, porém, que essa cadeia causal não é simples, mas surge a partir de uma complexa rede que ele denomina de *sobredeterminação*. Portanto, todo evento psíquico é causado por outros eventos anteriores; essa causalidade, porém,

é complexa e sua elucidação não se resolve numa simples regra de causa e efeito.

Em função dessa noção de sobredeterminação, a interpretação dos sonhos – e por extensão, a interpretação de cadeias associativas produzidas no decorrer do tratamento psicanalítico – trata em pé de igualdade os diversos elementos que surgem no relato do conteúdo manifesto e nas associações que se desprendem de seus elementos. Um elemento aparentemente pouco importante pode abrir caminho a cadeias associativas cruciais e que permitem melhor situar o material onírico na experiência psíquica do sujeito. Freud (2001, p. 496) insiste que “ao interpretar sonhos, atribuímos idêntica importância a cada um dos matizes de expressão linguística em que eles nos foram apresentados”, o que significa que mesmo temas caros à teoria analítica – como a imagem materna, representações eróticas ou objetos das pulsões – não devem ser tomados *a priori* como mais importantes que outros elementos do sonho.

O método de interpretação dos sonhos se pauta pela produção de cadeias associativas a partir dos diferentes elementos que compõem o relato do sonho. Tomando aspectos isolados do relato do sonho como ponto de partida, o sonhador deve associar livremente e falar sobre os pensamentos aos quais aquele elemento o remete. Esse processo é repetido para todos os elementos do relato, produzindo uma densa rede de associações que apresentam tanto pontas isoladas, pouco associadas a outros elementos, quanto nós que cruzam diferentes cadeias.

Mas o que permite à interpretação dos sonhos dar tamanho valor a detalhes que parecem acessórios e que, muitas vezes, podem ter sido acrescentados ao relato do sonho posteriormente? Um primeiro ponto, que afasta as críticas que assumem não ser o sonho interpretável por não termos acesso direto a ele, é o fato de que o relato do sonho não é considerado apenas uma formação sobre a experiência onírica, mas uma extensão do próprio processo onírico no nível de elaboração secundária (Freud, 2001). A deformação que o sonho sofre no relato, portanto, é uma extensão da deformação pela qual os pensamentos oníricos passam durante o trabalho do sonho. As alterações não são arbitrárias, mas motivadas pela mesma censura que produziu a experiência onírica original, e “estão associativamente ligadas ao material que substituem e servem para indicar-nos o caminho para esse material, que, por sua vez, pode ser o substituto de alguma outra coisa” (Freud, 2001, p. 497). As substituições que ocorrem durante o relato do sonho, tributá-



rias à elaboração secundária e às determinações conscientes, nada mais são do que novas modificações de um material que emergiu já como fruto de um processo de censura. Como essas substituições estão necessariamente ligadas ao material substituído por cadeias associativas, é possível recuperá-lo.

A continuidade entre relato do sonho e pensamentos latentes deixa evidente que por trás da *planificação* – no sentido do *aplatissement* de Laplanche (1980), ou seja, desconsiderando qualquer informação *a priori* sobre quais elementos seriam mais relevantes – das associações que emergem a partir dos elementos do conteúdo manifesto do sonho está uma noção de *simetria causal*. Ao mesmo tempo em que o conteúdo manifesto é causado pelo trabalho do sonho sobre os pensamentos oníricos, Freud pressupõe que os elementos do relato do sonho permitem causar a rememoração dos caminhos trilhados pelo trabalho do sonho. Sem essa simetria, as associações poderiam ser consideradas como fortuitas e pouco informativas. A importância dessa noção leva Freud a formalizá-la sob o conceito de *representações-meta*, que abordaremos logo abaixo, na Subseção 1.3.2.

A referência que Freud faz às transformações no relato do sonho serve de pista para indicar um aspecto importante das cadeias associativas. Se o que o sujeito associa ao falar sobre o sonho é uma extensão do processo de censura, sua fala está tão condicionada à força do recalque quanto o sonho estava no momento de sua produção. As associações, tal qual o relato do sonho, acabam por reunir duas exigências contrárias: tratam de abordar representações que se encontram no limite entre o inconsciente e o consciente, mas ao mesmo tempo não lhe é permitido avançar mais do que é cedido pelo processo de censura. Disso só é possível tirar uma conclusão: as associações livres são um tipo de formação de compromisso, um derivado do inconsciente adulterado pela censura – tal qual o próprio sonho, os atos falhos, os chistes e os sintomas. Na terceira seção das *Cinco lições de psicanálise*, Freud (1979) argumenta que, a despeito das dificuldades trazidas pelo abandono da hipnose e pela constatação de que nem sempre o primeiro elemento lembrado frente a uma hesitação das recordações era o elementos procurado, as associações deveriam necessariamente se parecer com o material recalcado, tal qual o sintoma. Um pensamento associado, em função da resistência, representa o recalcado de uma maneira indireta, sob a forma de *alusões*.

A estratégia proposta por Freud para resolver o problema da

validade do relato do sonho nos interessa particularmente. Quando confrontado com um relato confuso e de difícil compreensão, Freud pede para seu paciente relatar o sonho novamente. Apesar de os elementos da narrativa permanecerem similares, as palavras utilizadas e detalhes que a coloreem apresentam alguma variação. Nessa variação, Freud reconhece o trabalho da censura tentando encobrir possíveis detalhes reveladores no relato original (essa recomendação técnica aparece primeiramente em *A interpretação dos sonhos*, e é retomada mais tarde no escrito sobre *A psicanálise e a determinação dos fatos nos processos jurídicos*). O ponto que dessa estratégia nos interessa é que, para poder reconhecer quais elementos foram alterados ou sumiram do relato original, é necessário comparar, mesmo que intuitivamente, aspectos distribucionais dos termos utilizados para construir o relato. Identificar se um termo foi modificado, suprimido ou acrescentado implica em comparar, dentro de um contexto que aparece como comum entre os dois relatos, se a coocorrências dos termos permanece suficientemente semelhante ou se um elemento foi substituído por outro. Essa identificação só se torna possível se assumimos que, dentro de uma determinada sequência de termos, esperamos um termo com maior probabilidade do que outros. Se o termo esperado não é encontrado, mas é substituído por outro, podemos assumir que o novo termo apresenta alguma relação semântica com o termo suprimido – é seu substituto na formação de compromisso.

A substituíbilidade de um termo por outro dá uma ideia mais clara do que implica a dimensão semântica dos processos inconscientes para a Psicanálise. Diferente do entendimento de que o trabalho psicanalítico é eminentemente hermenêutico – como Grünbaum (1985) critica nos comentários de Habermas e Ricouer – a maneira como Freud aborda o funcionamento do aparelho psíquico na superfície das associações reflete muito mais um *mecanismo causal* entre representações conscientes e inconscientes do que uma técnica de geração de sentidos. O mecanismo causal que melhor representa essa característica do psiquismo é formalizada sob o conceito abrangente de *processo primário*. No processo primário, a energia psíquica é livre e pode se deslocar com facilidade entre diferentes representações, especialmente por meio dos fenômenos da condensação – quando aspectos de diversas representações são acumulados em uma representação que faz alusão às representações condensadas – e do deslocamento – quando a intensidade de uma representação é repassada para outra representação próxima, mas mais aceitável para a

censura (Freud, 2001). A liberdade da energia psíquica no processo primário diz respeito à tendência do psiquismo em descarregar uma excitação da maneira mais rápida possível por meio da identidade perceptiva, ou seja, reinvestindo o traço mnêmico ligado à experiência de satisfação. O processo secundário, que se desenvolve como uma elaboração posterior, força o excitação por um caminho indireto, objetivando a identidade de pensamento – em outras palavras, obter, pela via motora, uma experiência semelhante à experiência original. Um exemplo clássico de como o processo primário permite estabelecer uma identidade semântica entre diferentes termos é dado pela *equação simbólica*: “nos produtos do inconsciente [...] os conceitos de fezes (dinheiro, dádiva), bebê e pênis mal se distinguem um do outro e são facilmente intercambiáveis” (Freud, 1984a, p. 154). Assim como na repetição do relato do sonho, é a *substituibilidade* entre termos distintos na consciência que permitem inferir a existência de um núcleo semântico comum inconsciente.

O conselho prático de Freud e a noção de processo primário escondem, portanto, uma concepção distribucional do funcionamento do psiquismo. A hipótese distribucional (Sahlgren, 2006) é uma abordagem para lidar com a dimensão semântica da linguagem a partir do uso concreto da língua, retirando da repetição de padrões nos contextos onde aparece uma palavra as suas relações semânticas com outros termos. A intuição de Freud aqui segue o mesmo caminho: pela maneira como o analisando costuma organizar suas associações, é possível detectar quando um determinado termo ou locução é substituído por outros; e, dessa substituição, torna-se possível reconstruir as relações latentes entre as representações. Freud, porém, mantém-se fiel a seu axioma do absoluto determinismo psíquico; a hipótese distribucional, por sua vez, leva em consideração variações aleatórias nas substituições, focando-se nas relações semânticas em termos de distribuições de probabilidade. Apesar de não considerar o problema da incerteza na substituição dos termos, essa intuição distribucional freudiana servirá como fundamento para nosso modelo digital das associações livres. Antes, porém, precisamos discutir alguns modelos analógicos propostos por Freud.

### 1.3.1 Modelo geológico e modelo têxtil

A maneira como a interpretação dos sonhos reconstrói a continuidade da experiência psíquica a partir do relato do sonho é explicada pela conjugação de um modelo geológico e um modelo têxtil. Am-

bos os modelos usam de recurso à analogia para compreender os processos latentes que correspondem ao que o psicanalista escuta no tratamento, e servem como exemplo da criatividade de Freud em elaborar alegorias para facilitar a ilustração de seu entendimento sobre os fenômenos.

O modelo geológico do aparelho psíquico concebe a experiência consciente e as cadeias associativas inconscientes como diferentes camadas, com variáveis graus de dureza à penetração radial (Breuer & Freud, 1984). O processo de associação funciona como uma escavação de um sítio geológico, em que as camadas mais superficiais representam eventos cronológicos mais recentes e as camadas mais profundas estão associadas a experiências primeiras, como os traumas e as experiências eróticas infantis. Quanto mais o processo de investigação se aprofunda na exploração dos estratos, maior a resistência. O modelo geológico é muitas vezes complementado por uma variação arqueológica (Freud, 1984c), que permite estender a analogia para a decifração de traços descontínuos cujo significado e totalidade se perderam no tempo.

Em conjunto com o modelo geológico, Freud também recorre a um modelo têxtil. Um pouco menos visual que o modelo geológico, o modelo têxtil concebe o aparelho psíquico como formado por diversos fios associativos, com inúmeros pontos de enodamento ligando-os uns aos outros. Esses fios, por sua vez, são concebidos como combinações lineares de palavras. Essa linearidade é tributária da dimensão sintagmática da fala: durante o processo de enunciação, uma palavra sempre segue outra de maneira linear, formando uma cadeia cujos elos são dados pelas regras sintáticas da língua. Quando um determinado fio associativo é pinçado, por exemplo, a partir de um elemento do relato do sonho, o processo de livre associação tenderá a seguir essas cadeias associativas, traçando os caminhos que oferecem menor resistência – em outras palavras, caminhos que podem servir de compromisso entre as representações inconscientes recalçadas, impossibilitadas de emergir diretamente mas que pulsam para alcançar a consciência, e as exigências do recalque. No extremo, no curso da interpretação de um sonho, esse enodamento pode ser impossível de desfazer, formando um emaranhado que Freud denomina de *umbigo do sonho*.

### 1.3.2 Representações-meta e associacionismo

Em função da noção de sobredeterminação, Freud não concebe como possível a existência de qualquer atividade psíquica, consciente ou inconsciente, que seja uma perturbação puramente aleatória. Nos processos conscientes, essa determinação é mais ou menos evidente: toda atividade dos pensamentos tem um objetivo a ser alcançado. Esse objetivo é designado no psiquismo por uma representação especial, na medida em que governa o conjunto de ações mentais realizadas e as direciona para si, e recebe o nome de *representação-meta*. Por exemplo, no uso do raciocínio para resolver um problema de aritmética, os processos conscientes são conduzidos por uma representação-meta que é a resolução do problema. Porém, se não estamos engajados em uma atividade consciente com uma representação-meta explícita, estaria o psiquismo livre de qualquer objetivo?

Freud (2001) responde, categoricamente, que não. Apesar de só conseguir sustentar essa resposta com alguns poucos dados pontuais retirados do tratamento da neurose e da observação de fenômenos psicóticos, ele insiste que, quando abandonamos as representações-meta que dirigem a atividade consciente e, a partir de elementos do sonho, enunciamos todos os pensamentos que emergem involuntariamente na consciência, a atividade psíquica é guiada por representações-meta inconscientes ou, pelo menos, desconhecidas. Mas como esse processo de livre associação pode garantir que, ao longo das cadeias associativas, seja possível recuperar os pensamentos oníricos latentes? Um crítico poderia afirmar que a convergência das diversas cadeias associativas seria facilmente explicada pelo fato de que, tendo a primeira cadeia trazido determinados elementos à consciência, as cadeias que partem de outros elementos do relato do sonho irão se aproveitar das memórias já recordadas para retomá-las.

A solução encontrada por Freud é interessante, ainda que sua justificativa seja precária. Quando abandonamos as representações-meta conscientes, o processo de emergência de representações é dominado por representações com metas desconhecidas, determinando as associações que parecem involuntárias. A essa ideia ele opõe o associacionismo clássico, no qual uma representação que emerge o faz em resposta a uma representação provocadora. Não haveria apenas representações emergentes em função de reações, mas um empuxo de representações-meta desconhecidas para fazer com que determinadas cadeias associativas invadam a consciência. Uma cadeia de associações, portanto, não caminha aleatoriamente pelos complexos

que formam o psiquismo. Ela é determinada por representações fortemente investidas que exercem atração sobre outras representações. A ligação das cadeias com a representação-meta pode parecer fortuita num primeiro nível: assonância, similaridade, contiguidade. Mas o processo contínuo de associação revela que essa semelhança é facilitadora e substitui outra relação semântica mais profunda, em semelhança com a noção de *facilitação orgânica* ou *complacência somática*, utilizada como explicação da ligação entre os aspectos somáticos e psíquicos no sintoma de conversão histérica, e o papel dos restos diurnos no trabalho onírico. Nos três casos, uma relação fortuita entre um determinante físico e outro inconsciente permite ao processo primário fazer uso do determinante aparentemente não relacionado como um caminho para realização disfarçada de desejos recalçados.

Freud fundamenta nessa ligação mais profunda o que chamamos anteriormente de *simetria causal*: os pensamentos oníricos, passando pelo processo do trabalho do sonho, *causam* o sonho e, em extensão, o seu relato; ao mesmo tempo, associar livremente a partir de elementos retirados do relato do sonho *causa* a organização das associações a partir da representação-meta que deu origem ao sonho. A fragilidade do argumento se deve ao fato de que Freud não oferece, em *A interpretação dos sonhos*, evidências empíricas de que os processos associativos estariam de fato sendo determinadas por essas relações mais profundas, em vez de estarem sendo afetadas apenas por relações superficiais e fortuitas. Seu argumento é baseado em sua autoridade, exclusivamente, na medida em que invoca observações feitas com pacientes neuróticos e observações com pacientes psicóticos.

### 1.3.3 O conceito de complexo

Essa evidência surgirá mais tarde, quando C. G. Jung (1919) apresentar seus estudos feitos sobre os testes de associação de palavra, tomados como corroboração experimental definitiva para Freud. Em seu estudo, Jung faz uso do termo *complexo* para se referir à constelação de termos associados entre si, marcadas por um forte colorido emocional, que podem emergir em função de uma palavra-estímulo. A proposição de existência de termos associados e emocionalmente marcados se mostrou necessário para justificar a interveniência, nas respostas a palavras-estímulos semanticamente distintas, de respostas intimamente relacionadas entre si. Em particular, era notável

que quando um complexo em particular dominava a atividade psíquica do sujeito, suas respostas tendiam a refletir muito mais o conteúdo do complexo do que as palavras-estímulo empregadas. Mesmo assim, a noção de complexo nunca é formulada de maneira apropriada, e seu sentido é tomado como relativamente evidente. É apenas numa curta nota de rodapé que C. G. Jung (1919, p. 66) dará uma definição mais completa: “por *complexo colorido por afetos* compreendemos o conjunto de representações relacionadas a uma experiência definida que é carregada de emoção. No que segue, sempre utilizaremos o termo *complexo* com esse sentido”.

Influenciado pelo termo utilizado pela escola de Zurique, Freud (1984f) chega a fazer uso do termo *complexo* para se referir à fonte da qual derivam as associações no lugar da representação-meta. A substituição não é gratuita, pois Freud aproxima as descobertas provenientes da clínica psicanalítica à descoberta de que as palavras-reação num teste de associação de palavras podem ser atribuídas a um conjunto de ideias com colorido emocional. Todavia, Freud já havia feito uso do termo complexo antes da influência dos estudos de Jung, apesar de renegar a autoria do conceito e atribuí-lo exclusivamente à escola de Zurique. Esse uso aparece em *Estudos sobre a Histeria* (Breuer & Freud, 1984), para designar uma representação intuitiva do modo como redes associativas estão dispostas de maneira mais ou menos fixas, guiando a ordem das associações. A utilização precoce do termo complexo demonstra a relevância, na representação de Freud sobre o funcionamento do psiquismo, da noção de que representações se organizam em conjuntos que compartilham algum tipo de propriedade em comum.

Seja utilizando como referência o conceito de complexo ou de representação-meta, as associações são consideradas livres mediante algumas ressalvas (Laplanche & Pontalis, 1998):

1. Apesar de haver elementos eliciadores das associações – como as palavras-estímulo ou o relato do conteúdo manifesto do sonho – seu desenvolvimento é livre porque não estão condicionadas a uma intenção seletiva consciente, ou, mais especificamente, a uma representação-meta ou complexo consciente;
2. A liberdade, portanto, diz respeito a coibir qualquer intenção seletiva consciente, o que não significa que as associações são completamente indeterminadas. Uma vez abandonadas as representações-meta conscientes, as cadeias se dirigem a representações-meta ou complexos associativos inconscientes.

Freud deixa claro que a maneira como concebe a relação entre representações no psiquismo é tributária, como vimos, do associacionismo (Freud, 2001). pelo qual foi fortemente influenciado durante sua formação, uma vez que esse paradigma de compreensão dos fenômenos psíquicos se encontrava em alta nos países germânicos durante o séc. XIX (Laplanche & Pontalis, 1998). Para o associacionismo, uma associação é, em suma, uma ligação entre elementos psíquicos, mas também é utilizado, por metonímia, para designar elementos associados entre si. A série de ligações entre diferentes elementos psíquicos é denominada de cadeia associativa.

O uso que Freud faz dos termos associacionistas, todavia, foi transformado por sua experiência clínica. A ideia de associação como cadeias de palavras produzidas a partir do método de tratamento é original da Psicanálise (Laplanche & Pontalis, 1998). A noção de que a cadeia de associações obedece a representações-meta também marca o rompimento de Freud com o associacionismo clássico, para o qual os pensamentos se associam reativamente por contiguidade e semelhança; para Freud, as representações-meta estabelecem relações mais profundas e significativas entre elementos associados, tornando desnecessário o pressuposto de que uma reação deve partir de um elemento eliciador fortuito.

### 1.3.4 A necessidade de vários registros do psiquismo

As associações verbalizadas pelos pacientes em análise são denominadas por Freud de *Einfall*. Seu sentido mais imediato para o processo clínico é de uma ideia súbita, um pensamento que emerge de maneira independente da vontade consciente. Entretanto, o termo também implica a invasão de um território estrangeiro: uma imagem que corresponde à topologia de diferentes registros psíquicos que se influenciam mutuamente (Laplanche & Pontalis, 1998).

Um aspecto crucial de uma associação, como uma “ideia que vem à cabeça”, é que, mesmo parecendo isolada, remete sempre a outras ideias, graças às representações-meta inconscientes. O que surge, em verdade, são séries associativas denominadas por Freud de diversos nomes: cadeias, linhas, trens. Essas séries se encontram em diversos pontos, formando não só cadeias unidimensionais, mas redes com dimensões múltiplas – um aspecto que fica evidente no modelo têxtil descrito acima, na seção Subseção 1.3.1. O encadeamento tomado pelas associações segue uma organização complexa da memória, que ele representa como análoga a um arquivo com-



posto por diversos sistemas de classificação: os traços mnêmicos de um mesmo evento podem estar representados em diferentes sistemas de memória.

Nesses sistemas mnêmicos existem *complexos associativos* de representações ou traços mnêmicos. Parte desses complexos se encontra cindido, separado do percurso associativo. Alguns elementos da parte cindida dos complexos podem retornar, isoladamente, para a consciência, mas determinadas combinações dessas representações têm seu acesso interditado. Esses agrupamentos de representações separados sugerem a existência de diferentes tópicas do psiquismo e exigem a existência de um registro psíquico inconsciente, bem como a existência de uma força contrária à emergência dos elementos cindidos na cadeia associativa.

A operação da censura sobre essas representações opera por duas maneiras similares, aproveitando-se de facilitações de associações mais superficiais. Na primeira, representações que por si só não provocam objeção têm sua ligação problemática recalcada, mantendo-se vinculadas nas cadeias associativas apenas por outras ligações mais tênues e contingenciais. Na segunda, duas representações são recalcadas em função de seu conteúdo e suas representações substitutivas são escolhidas com base numa associação que reproduza, de maneira superficial, a relação mais profunda dos pensamentos censurados (Freud, 2001).

A substituição das representações-meta conscientes por representações-meta inconscientes, e o processo pelo qual as associações superficiais entre representações são utilizadas como substitutos de relações mais profundas entre representações recalcadas são descritos por Freud como “*teoremas* [que se transformaram] em pilares básicos da técnica psicanalítica” (Freud, 2001, p.513, grifo nosso). Freud acrescenta ainda outra representação-meta fundamental ao tratamento: a representação que está relacionada ao analista, aspecto que será desenvolvido posteriormente em sua obra como o problema da *transferência*.

Freud reconhece que a fundamentação teórica do processo de livre associação é condição essencial para compreender a prática analítica. De fato, se os dados obtidos a partir das cadeias associativas não tiverem relações claras com o aspecto inconsciente dos complexos de representações, nenhuma inferência sobre o inconsciente será possível, e voltamos ao problema de que o analista pode ser apenas um leitor de pensamentos. Chegamos aqui a um ponto crucial: fundamentar a técnica de associação livre é, portanto, estabelecer uma

ponte entre a prática e a teoria analítica. Nossa tarefa de construir um modelo digital para as associações livres se mostra não só uma exigência científica, como evidenciado pelas críticas à pesquisa em Psicanálise abordadas na Introdução, mas uma tarefa fundamental à Psicanálise independente de qualquer outra demanda externa, como esperamos ter demonstrado retomando as elaborações freudianas sobre o tema. Antes de iniciarmos a construção de nossa proposta, porém, resta ainda abordar em profundidade os modelos probabilísticos que nos servirão de base: os *modelos de tópicos*.

## Capítulo 2

# Modelos de tópicos

Se desejamos construir modelos digitais para os fenômenos sobre os quais pretendemos aprender algo a respeito, é crucial escolher com cuidado quais objetos matemáticos representarão os elementos essenciais do sistema em questão. No nosso caso, estamos interessados em elaborar uma representação formal do processo de livre associação. Para passarmos dos modelos analógicos propostos por Freud para modelos digitais que consigam reproduzir aspectos essenciais de como a livre associação dá acesso a representações inconscientes, teremos de fazer escolhas para abstrair características importantes e construir um modelo idealizado.

Apresentamos na Introdução a possibilidade de este modelo formal ser tomado de empréstimo dos desenvolvimentos recentes na área de processamento de linguagem natural por computador. Em particular, interessa-nos os avanços dos modelos de representação semântica chamados de *modelos de tópicos*. Esses modelos podem ser considerados como algoritmos de aprendizagem de máquinas não supervisionados, ou seja, capazes de descobrir padrões em um conjunto de dados a partir de suas estruturas latentes e de um conjunto de pressupostos. No caso dos modelos de tópicos, essas estruturas latentes são conjuntos temáticos – os *tópicos* – que reúnem palavras semanticamente próximas e que determinam quais palavras são escolhidas na construção de um documento, de acordo com um modelo idealizado de como documentos são produzidos.

Porém, não queremos apenas aplicar algoritmos emprestados dos modelos de tópicos como uma ferramenta alheia à teoria da Psicanálise, um simples instrumento para auxiliar na investigação empírica. Nossa tese central é, justamente, que alguns aspectos da teoria psicanalítica ligados ao processo de livre associação podem ser representados como modelos formais. Por esse motivo, é crucial compreender não apenas o núcleo duro dos modelos de tópicos, de forma a permitir entender seus resultados, mas também toda rede

conceitual que os fundamenta.

David Blei, um dos autores do artigo seminal sobre modelos de tópicos (Blei, Ng, & Jordan, 2003), argumenta que a *Alocação de Dirichlet Latente* – o primeiro modelo de tópico de utilização mais ampla, base do modelo que utilizaremos nesta tese – nada mais é do que um estudo de caso de inferência bayesiana para modelos gráficos probabilísticos que representam processos gerativos em modelos de mistura. Para instrumentar nossa compreensão dos modelos de tópicos, portanto, precisamos conhecer primeiramente como funciona a inferência bayesiana; o que são modelos gerativos e como eles podem ser representados por grafos; como é possível computar, de modo aproximado, a inferência quanto utilizamos esses modelos. A exposição desses conceitos é feita com amplo apoio na obra *Bayesian Data Analysis* (Gelman, Carlin, Stern, & Rubin, 2014), um livro reconhecido pela sua abordagem aprofundada da inferência bayesiana, e também nas obras derivadas como *Doing Bayesian Data Analysis* (Kruschke, 2010) e *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists* (Lynch, 2007), livros de caráter mais aplicado voltado para pesquisadores em Ciências Sociais e Humanas. Sustentados nesses conceitos poderemos, finalmente, abordar os modelos de tópicos.

## 2.1 Inferência bayesiana

A teoria da probabilidade surgiu na matemática de uma maneira pouco usual. Inicialmente formulada para analisar jogos de azar, sua influência nas diversas ciências se tornou cada vez mais patente na medida em que permite formalizar a incerteza inerente a todo processo de inferência científica. Uma vez que na construção do conhecimento científico nunca temos acesso a informações completas e perfeitas – seja por erros de mensuração, inacessibilidade dos fenômenos à observação direta ou por amostras limitadas – a testagem de hipóteses e a formulação de teorias precisam levar em consideração as diversas fontes de incerteza sobre os fenômenos estudados. A teoria da probabilidade oferece ferramentas para operacionalizar essa tarefa (Griffiths, Kemp, & Tenenbaum, 2008).

Dentre essas ferramentas, a inferência bayesiana e seu principal fundamento, a regra de Bayes, são desenvolvimentos da teoria da probabilidade para resolver o problema da probabilidade inversa. Esse problema reflete o que um cientista encontra na produção de conhecimento: tendo acesso a um conjunto limitado de dados ob-

servados, o que é possível concluir sobre o processo que gerou esses dados? Caracterizar como *inversa* a formulação dessa classe de problemas se deve à maneira como a teoria da probabilidade clássica partia do pressuposto de que os parâmetros e o processo de geração de dados eram conhecidos e passava, então, a quantificar a probabilidade do acontecimento de determinados eventos (Fienberg, 2006).

### 2.1.1 Regra de Bayes

A regra de Bayes é um resultado elementar da aplicação dos axiomas básicos da probabilidade, e sua descoberta é atribuída ao Reverendo Thomas Bayes, ainda que a aplicação do processo de inferência de probabilidade inversa tenha sido generalizado por Pierre Simon Laplace (Fienberg, 2006), ambos no séc. XVIII. Suponhamos que temos duas variáveis aleatórias,  $A$  e  $B$ . Denotamos a probabilidade marginal de qualquer valor de  $A$  como  $p(A = a)$  ou, de maneira mais simples,  $p(a)$ ; da mesma forma, a probabilidade marginal de  $B$  é denotada como  $p(B = b)$  ou simplesmente  $p(b)$ . A *regra da cadeia* na teoria da probabilidade permite calcular a probabilidade conjunta das duas variáveis apresentarem determinados valores. A probabilidade conjunta é denotada pela expressão  $p(A = a \cap B = b)$ , normalmente simplificada para  $p(a, b)$ . Pela regra da cadeia, a probabilidade conjunta é dada pelo produto da probabilidade marginal de uma variável com a probabilidade condicional da outra variável, dado o valor assumido para a variável anterior.

$$p(a, b) = p(a)p(b | a) \quad (2.1)$$

Ou, de maneira menos parcimoniosa,  $p(A = a \cap B = b) = p(A = a)p(B = b | A = a)$ . Como  $p(a, b) = p(b, a)$ , podemos fatorar a distribuição conjunta utilizando a distribuição marginal de  $B$  e a distribuição condicional de  $A$  dado  $B$ . Disso decorre que:

$$p(a)p(b | a) = p(b)p(a | b) \quad (2.2)$$

$$p(b | a) = \frac{p(b)p(a | b)}{p(a)} \quad (2.3)$$

$$\text{posterior} = \frac{\text{priori} \cdot \text{verossimilhança}}{\text{evidência}} \quad (2.4)$$

A Equação 2.3, acima, é a regra de Bayes, derivada diretamente da regra da cadeia (Kruschke, 2010). Sua característica mais

marcante é permitir computar a probabilidade de  $B$  condicionada a valores dados para  $A$  a partir da probabilidade condicional de  $A$  dado  $B$ .

A regra de Bayes se torna particularmente útil se assumimos determinados pressupostos para as variáveis que nos interessam e interpretamos a probabilidade não como frequência assintótica, mas como graus de crença. Por exemplo: um cientista obteve, por meio de um experimento, um conjunto de dados  $d$ . Ele tem uma série de hipóteses  $h$  sobre o processo que gerou os dados observados. Antes de observar os dados, o cientista assume valores de probabilidade sobre qual hipótese é verdadeira. A probabilidade  $p(h)$  que o cientista dá para as várias hipóteses é chamada de *probabilidade a priori*, e indica sua crença prévia em quais hipóteses são mais ou menos prováveis<sup>1</sup>. O problema que ele enfrenta, tendo obtido determinado conjunto de dados, é: como deve atualizar suas crenças com base nos dados observados? O que ele quer é calcular o valor de  $p(h | d)$ , ou seja, a probabilidade de uma hipótese ser verdadeira condicionada aos novos dados disponíveis. Esse valor é chamado de *probabilidade posterior* ou *a posteriori*. Para calcular a probabilidade posterior, o cientista pode utilizar a regra de Bayes:

$$p(h | d) = \frac{p(h)p(d | h)}{p(d)} \quad (2.5)$$

Ou seja, a probabilidade posterior é proporcional à probabilidade *a priori* e a probabilidade dos dados condicionada à hipótese –  $p(d | h)$ . Esse termo é denominado de *verossimilhança* e constitui um elo crucial na inferência bayesiana, uma vez que ele indica sob qual hipótese os dados observados são mais verossímeis. O termo  $p(d)$ , ou seja, a probabilidade marginal dos dados, é denominado de *evidência* ou *verossimilhança marginal*. Utilizando outra regra da probabilidade, a marginalização ou regra da soma, a probabilidade da evidência pode ser calculada como a soma da distribuição conjunta das duas variáveis com relação a todos os valores de uma das variáveis. Por exemplo,  $p(b) = \sum_a p(a, b)$ , ou seja, a probabilidade marginal de  $B = b$  pode ser obtida somando a probabilidade conjunta de  $B = b$  para todos os valores possíveis de  $A$ . Com isso, a regra de Bayes pode ser reescrita da seguinte maneira:

<sup>1</sup>Dentro do referencial frequentista, outra perspectiva de interpretação da probabilidade e de inferência estatística, atribuir valores de probabilidade a hipótese ou parâmetros não faz sentido, porque eles são considerados fixos e não relacionados ao desfecho assintótico de uma série de ensaios dentro de uma mesma classe de referência.

$$p(h \mid d) = \frac{p(h)p(d \mid h)}{\sum_{h' \in \mathcal{H}} p(h')p(d \mid h')} \quad (2.6)$$

A soma no denominador da regra de Bayes garante que a distribuição posterior é uma distribuição de probabilidade própria, ou seja, a soma de todos os valores possíveis é igual a um.

### 2.1.2 Regra de Bayes aplicada à testagem de hipóteses

A regra de Bayes pode ser aplicada para comparar a força de evidência apresentada por um conjunto de dados em favor de uma hipótese. A maneira mais simples de ilustrar seu uso nesse contexto é apresentar a derivação da regra de Bayes para quando o número de hipóteses é igual a dois. O primeiro passo necessário é estabelecer o espaço de hipóteses  $\mathcal{H}$ . No caso de duas hipóteses, esse espaço pode ser representado como dois valores pontuais para um parâmetro de interesse. Estabelecidas as hipóteses, é necessário definir a probabilidade *a priori* de cada uma delas –  $p(h)$  – bem como a verossimilhança dos dados obtidos sobre cada hipótese considerada –  $p(d \mid h)$ .

A regra de Bayes pode ser então utilizada para computar o valor da distribuição posterior para ambas as hipóteses,  $p(h_1 \mid d)$  e  $p(h_2 \mid d)$  (Griffiths et al., 2008). Porém, no caso de haver efetivamente apenas duas hipóteses, podemos compará-las diretamente calculando a chance posterior, que nada mais é do que a razão (*odds*) das duas distribuições posteriores.

$$\frac{p(h_1 \mid d)}{p(h_2 \mid d)} = \frac{p(h_1)p(d \mid h_1)}{p(h_2)p(d \mid h_2)} \quad (2.7)$$

O termo referente à evidência, tal qual apareceu nas equações anteriores, é cancelado nessa formulação por ser o mesmo para as duas hipóteses. Sob essa formulação, o primeiro termo do lado direito da equação é chamado de *chances a priori*, e indica a razão da crença prévia à observação dos dados sobre qual hipótese é mais provável. O segundo termo é denominado de *razão de verossimilhança* e indica sob qual hipótese os dados são mais prováveis. Na formulação acima, com  $h_1$  no numerador, o resultado da equação é a chance posterior de  $h_1$  em relação a  $h_2$ , que indica quanto a evidência dada pela razão de verossimilhança modifica a razão *a priori* entre as hipóteses. O valor da razão de verossimilhança também

é denominado de *fator de Bayes* quando a comparação entre dois modelos é baseada na probabilidade da evidência sob cada um dos modelos avaliados, e é sugerido como um substituto ao uso de valores-p para testagem de hipóteses (Kruschke, 2010). É conveniente notar que a chance posterior é igual ao fator de Bayes quanto a probabilidade *a priori* das duas hipóteses é igual.

A regra de Bayes facilita a comparação entre modelos com diferentes graus de complexidade, tornando a escolha de modelos uma extensão intuitiva do quadro de referência já utilizado para testar hipóteses e estimar parâmetros. Uma característica interessante da seleção de modelos bayesiana é que a complexidade do modelo é automaticamente levada em consideração quando se comparam duas hipóteses de diferentes complexidades. Uma hipótese mais restritiva não terá a mesma flexibilidade que uma hipótese com maior número de graus de liberdade para ajustar os dados obtidos, o que a princípio favorece hipóteses mais flexíveis. Porém, essa flexibilidade também permite que ajustes piores existam, e a marginalização com relação aos parâmetros do modelo ameniza a relação entre perdas e ganhos de hipóteses com diferença de complexidade. Essa penalização automática dada a modelos mais flexíveis na inferência bayesiana é denominada de “Navalha de Occam bayesiana” (Griffiths et al., 2008).

### 2.1.3 Estimativa de parâmetros

Nem sempre uma pergunta de pesquisa consegue reduzir sua estimativa de interesse à razão entre duas hipóteses concorrentes, mutuamente exclusivas e exaustivas. Pelo contrário, costuma ser mais comum numa investigação científica buscar o valor mais apropriado para um parâmetro de interesse a partir de um espaço de hipótese infinito. Nesse caso, o problema de inferência não é apenas uma comparação entre um número finito de hipóteses, mas a *estimativa de parâmetros* (Griffiths et al., 2008). O *problema da estimativa*, entendido como o cálculo dos valores mais prováveis dos parâmetros do modelo concebido como aquele que gerou um grupo de dados, em conjunto com o *problema da predição*, compreendido como a inferência da probabilidade de novas observações em função das observações já feitas, podem ser considerados como os principais objetivos do processo de inferência estatística (Heinrich, 2009), além do problema já abordado do uso de dados para decidir entre um conjunto finito de modelos (Kruschke, 2010).



O caso do modelo utilizado nesta tese configura, ao mesmo tempo, um problema de estimativa e de predição. Uma vez que a formalização da Psicanálise é ainda incipiente, não será possível deduzir hipóteses pontuais ou mesmo intervalares. O modelo de tópicos, cuja descrição será dada em detalhes logo abaixo, faz uso da regra de Bayes para estimar os valores de variáveis latentes presumidas pelo modelo. Os valores dessas variáveis latentes serão obtidos com base em um conjunto de dados observados; neste caso, as palavras enunciadas ao longo de diferentes sessões. De posse do modelo ajustado, ou seja, com suas variáveis latentes já computadas, podemos utilizá-lo para prever novos eventos. Com isso, podemos avaliar quão bem o modelo prediz aspectos que nos interessam do processo de livre associação.

### 2.1.3.1 Estimativa por Máxima Verossimilhança

Na estatística frequentista<sup>2</sup>, o problema da estimativa costuma ser resolvido utilizando um estimador pontual, obtido a partir da aplicação da maximização ou minimização de uma função dos dados. Um dos estimadores mais utilizados é aquele que *maximiza a verossimilhança*, ou seja, a estimativa que torna os dados obtidos mais prováveis. Além disso, é comum também computar o erro padrão atribuído a esse estimador para obter uma medida da incerteza de seu valor (Gelman & Hill, 2006).

Se o modelo sobre o qual queremos realizar inferência é composto por um conjunto  $\theta$  de parâmetros desconhecidos, e obtemos do sistema em questão um conjunto de dados  $\mathcal{X}$  que trazem alguma informação sobre os parâmetros de interesse, podemos utilizar a *máxima verossimilhança* para computar os valores mais prováveis para os parâmetros. A função de verossimilhança é uma função de probabilidade dos dados observados (e portanto fixos) condicionada a valores variáveis do parâmetro de interesse. Ela pode ser definida da seguinte maneira, se assumimos que os dados são realizações independentes e identicamente distribuídas das variáveis aleatórias em estudo (Heinrich, 2009):

---

<sup>2</sup>Apesar da probabilidade ser uma teoria matemática bem fundamentada e com amplos desenvolvimentos, existem divergências quanto à interpretação do que exatamente significa probabilidade no mundo real. Duas tendências centrais são o bayesianismo, apresentado aqui e crucial para o desenvolvimento do modelo utilizado neste trabalho de tese, e o frequentismo, perspectiva mais conhecida baseada na interpretação da probabilidade como limite da frequência de ocorrência numa série infinita de ensaios dentro de uma mesma classe de equivalência.

$$L(\theta \mid \mathcal{X}) \triangleq p(\mathcal{X} \mid \theta) = \prod_{x \in \mathcal{X}} p(x \mid \theta) \quad (2.8)$$

ou seja, a probabilidade conjunta dos dados observados ocorrerem condicionada a diferentes valores do parâmetro  $\theta$ . O pressuposto de independência e identidade de distribuição simplifica o cálculo da distribuição conjunta dos dados observados: sob esse pressuposto, a verossimilhança total é simplesmente o produto da verossimilhança de cada ponto de dado. Como a função de verossimilhança resulta num produto, é usual substituí-la pela função do logaritmo da verossimilhança, o que torna sua computação mais tratável. A log-verossimilhança é denotada como  $\mathcal{L} \triangleq \log L$ . A maximização da verossimilhança é então definida como

$$\hat{\theta}_{ML} = \arg \max_{\theta} \mathcal{L}(\theta \mid \mathcal{X}) = \arg \max_{\theta} \sum_{x \in \mathcal{X}} \log p(x \mid \theta) \quad (2.9)$$

ou seja,  $\hat{\theta}_{ML}$  é o argumento que maximiza a função de verossimilhança e é o estimador pontual dos parâmetros de interesse. Esse valor é obtido a partir da resolução da primeira derivada de  $\mathcal{L}$  quando seu valor é igualado a zero, o que significa que o estimador é a moda da função de verossimilhança (Heinrich, 2009).

Um problema do estimador obtido por meio da máxima verossimilhança é a possibilidade de superajuste (*overfit*), o que o torna pouco confiável para generalização para novos casos. Outro problema é o fato de que o estimador obtido por máxima verossimilhança leva em consideração apenas o conjunto de dados obtido num ensaio, descartando qualquer informação prévia que não esteja incluída nesse conjunto de dados. Em função disso, existe a possibilidade dos parâmetros não serem *identificados* pelo conjunto de dados, tornando a estimativa pouco estável.

### 2.1.3.2 Estimativa por máximo *a posteriori* (MAP)

Uma alternativa à máxima verossimilhança é a estimativa por *máximo a posteriori* (MAP), que permite incluir informações prévias à análise dos dados observados. Isso é feito pelo acréscimo de uma distribuição *a priori*,  $p(\theta)$ , à função de verossimilhança.

$$\begin{aligned}
\hat{\theta}_{MAP} &= \arg \max_{\theta} p(\theta \mid \mathcal{X}) \\
&= \arg \max_{\theta} \frac{p(\theta)p(\mathcal{X} \mid \theta)}{p(\mathcal{X})} \\
&= \arg \max_{\theta} p(\theta)p(\mathcal{X} \mid \theta) \\
&= \arg \max_{\theta} \{\log p(\theta) + \mathcal{L}(\mathcal{X} \mid \theta)\} \\
&= \arg \max_{\theta} \left\{ \sum_{x \in \mathcal{X}} \log p(x \mid \theta) + \log p(\theta) \right\}
\end{aligned} \tag{2.10}$$

O acréscimo de uma distribuição *a priori* sobre os parâmetros de interesse abre caminho para ampliar o alcance do processo de estimativa. A distribuição *a priori* pode ser interpretada como o acréscimo de informações prévias sobre os parâmetros. Outra maneira de interpretar a distribuição *a priori* é como um termo de regularização da estimativa de máxima verossimilhança, como estratégia para evitar o superajuste ou permitir a identificação do parâmetro quando os dados são pouco informativos (Heinrich, 2009).

A estimativa por MAP faz uso da concepção bayesiana de probabilidade, o que implica pensar os parâmetros como variáveis aleatórias às quais é possível atribuir uma distribuição de probabilidade – um requisito desnecessário para estimação por máxima verossimilhança. É possível modelar essa distribuição de probabilidade *a priori* a partir de uma família paramétrica, ou seja, a distribuição *a priori* para os parâmetros é modelada a partir de outra série  $\alpha$  de hiperparâmetros,  $p(\theta \mid \alpha)$ , criando uma hierarquia de parâmetros.

Apesar do uso da regra de Bayes e da atribuição de distribuições de probabilidade aos parâmetros a serem estimados, o estimador  $\hat{\theta}_{MAP}$  sofre de algumas limitações. Primeiro, ele permanece apenas um estimador pontual que indica o valor de maior probabilidade na distribuição posterior. Apesar de ser possível estimar o erro padrão desse estimador, o MAP não oferece uma descrição completa da distribuição posterior. Outro limite é que o estimador MAP se refere à moda da distribuição posterior, o que significa que não pode ser considerado um estimador ótimo para funções de perda quadráticas ou absolutas.

### 2.1.3.3 Estimativa bayesiana

A inferência bayesiana estende a estimativa pelo método de MAP computando toda a distribuição de probabilidade do conjunto de parâmetros  $\theta$  em vez de fazer uso de um estimador pontual. Com isso, além do valor mais provável dos parâmetros condicionados aos dados e às informações prévias, a inferência bayesiana apresenta informações mais ricas sobre o processo de inferência, pois permite computar qualquer função sobre a distribuição posterior, como o valor esperado, a variância ou intervalos entre percentis. Como se trata de uma distribuição de probabilidade sobre os parâmetros, sua variância pode ser utilizada como uma medida da incerteza sobre seus possíveis valores.

No caso do problema da estimativa de um ou mais parâmetros, a regra de Bayes pode ser diretamente aplicada, mas dessa vez não mais utilizando apenas massas de probabilidade pontuais, mas sim funções de densidade de probabilidade. Tratando agora cada elemento da equação não mais como um número específico mas como funções, podemos reescrever a regra de Bayes da seguinte maneira:

$$p(\theta \mid \mathcal{X}) = \frac{p(\theta)p(\mathcal{X} \mid \theta)}{p(\mathcal{X})} \quad (2.11)$$

na qual  $\theta$  é o parâmetro ou vetor de parâmetros de interesse, composto por um número infinito de valores possíveis – mais especificamente, uma distribuição sobre esses valores – e os fatores presentes na equação recebem a mesma interpretação já apresentada acima na Equação 2.4.

Por meio do uso de distribuições de probabilidade *a priori*, a inferência bayesiana obtém pelo menos duas vantagens com relação à estimativa pela função de verossimilhança: ela consegue fazer uso de informações prévias às coletas de dados, como resultados de pesquisas anteriores, observações e intuições; e também pode evitar o superajuste das estimativas regularizando o valor do parâmetro inferido. Porém, como no processo de testagem de hipóteses, a escolha da distribuição *a priori* para os parâmetros modifica os resultados obtidos. Quando não há informação que torne um valor ou um intervalo de valores mais provável do que outros, é comum a utilização de distribuições *a priori* não-informativas ou fracamente informativas. Uma das distribuições mais utilizadas nesse caso é a distribuição uniforme, que atribui igual probabilidade a cada valor do espaço de parâmetros (Gelman & Hill, 2006). Seu uso, porém,

pode ser questionado de acordo com o contexto de uso do modelo: é possível que haja informações prévias relevantes que devem ser levadas em consideração na estimação dos parâmetros; ou, no caso do uso da distribuição uniforme para representar completa ausência de informação, é importante ter em mente que ela não é invariante a transformações dos parâmetros.

A distribuição posterior indica os valores mais prováveis para os parâmetros e a incerteza relativa a esses valores. Em função da riqueza de informações contidas na distribuição posterior, a análise de dados bayesiana geralmente não a reduz a um único número, preservando a incerteza no processo de inferência. Quando há interesse em resumir a distribuição posterior, os números costumeiramente escolhidos são a moda da distribuição, que corresponde ao estimador que maximiza a distribuição posterior; a mediana, quando a função de perda considerada para o estimador se baseia no erro absoluto; ou o valor esperado ou a média da distribuição, quando a função de perda se baseia no erro quadrático. Além, é claro, da possibilidade de utilizar a variância e a distância entre percentis como uma medida da incerteza sobre os valores dos parâmetros (Gelman et al., 2014).

Uma dificuldade para a realização da inferência bayesiana completa é devida ao cálculo do denominador da regra de Bayes. Para a inferência bayesiana ser levada a cabo, não basta otimizar a função de probabilidade posterior para encontrar o argumento que maximiza a probabilidade dos parâmetros, mas é necessário calcular toda a distribuição posterior utilizando a regra de Bayes. Isso torna necessário o cálculo do denominador, a *evidência*, que se torna uma integral quando o espaço dos valores para os parâmetros é contínuo:

$$p(\mathcal{X}) = \int_{\theta \in \Theta} p(\theta) p(\mathcal{X} | \theta) d\theta \quad (2.12)$$

O cálculo desse termo, necessário para a normalização da distribuição posterior, costuma ser intratável na maioria dos problemas aplicados da inferência bayesiana (Lynch, 2007).

### 2.1.4 Computando a distribuição posterior

A dificuldade em resolver analiticamente o denominador da regra de Bayes para problemas de estimativa levou à pesquisa de procedimentos que permitissem computar a distribuição posterior sem resolver explicitamente o problema do cálculo da evidência. Dentre as solu-

ções analíticas possíveis, abordaremos a estratégia do uso de *distribuições conjugadas*, que, apesar de oferecer uma solução elegante e de fácil resolução, tem aplicabilidade limitada. Frente à limitação do uso de soluções analíticas, surgiu uma série de métodos numéricos que permitem aproximar a distribuição posterior. Um dos métodos mais utilizados é a simulação numérica utilizando Monte Carlo por Cadeias de Markov (*Markov Chain Monte Carlo*, *MCMC*), que também será brevemente abordado abaixo, por ser uma das principais estratégias de estimação para modelos de tópicos. Outra estratégia comum no contexto de modelos para dados textuais é a inferência variacional, que também abordaremos brevemente por sua importância no algoritmo utilizado neste trabalho. Há, ainda, outras estratégias de aproximação que não abordaremos aqui, tais como a expectância-maximização (*expectation maximization*) e propagação de expectância (*expectation propagation*) (Y. Wang, 2008).

#### 2.1.4.1 Distribuições conjugadas

Uma das soluções para lidar com a intratabilidade do cálculo da evidência no processo de inferência bayesiana é a utilização de distribuições *a priori* conjugadas com a função de verossimilhança. Dependendo da função de verossimilhança utilizada para um determinado problema de estimativa de um parâmetro, é possível encontrar distribuições *a priori* que possuem uma propriedade interessante: a forma paramétrica da distribuição posterior segue a mesma forma da distribuição *a priori*. Essas distribuições são denominadas de *distribuições conjugadas* e facilitam a resolução analítica de problemas de inferência bayesiana (Gelman et al., 2014).

Os parâmetros da distribuição *a priori* são denominados de *hiperparâmetros*, e no caso de distribuições conjugadas, seus valores são geralmente interpretáveis no contexto dos dados coletados. Uma *distribuição conjugada*, então, é qualquer distribuição *a priori* que, quando utilizada com determinadas funções de verossimilhança, resulta numa distribuição posterior com a mesma forma funcional da distribuição *a priori* mas cuja parametrização incorpora os dados observados. Com isso, o termo de evidência da regra de Bayes não precisa ser diretamente calculado, estando embutida na forma da função de probabilidade utilizada para descrever as informações disponíveis *a priori* (Heinrich, 2009).

**Exemplo de distribuição conjugada para distribuições multinomiais** Tradicionalmente, a noção de conjugalidade é demonstrada utilizando um modelo baseado nas distribuições binomial e beta. Porém, esse modelo básico tem pouca utilidade para problemas mais complexos, como é o caso da análise de dados textuais. Se generalizarmos o número de resultados possíveis do modelo binomial de dois para  $K$ , obtemos uma generalização da distribuição de Bernoulli: a distribuição *multinomial* (Heinrich, 2009). Sua formalização é semelhante à definição da distribuição binomial:

$$p(\vec{n} \mid \vec{p}, N) = \binom{N}{\vec{n}} \prod_{k=1}^K p_k^{n^{(k)}} \triangleq \text{Mult}(\vec{n} \mid \vec{p}, N) \quad (2.13)$$

Com as restrições de que  $\sum_k p_k = 1$ , para que a distribuição seja própria; e  $\sum_k n^{(k)} = N$ , ou seja, a soma dos resultados de diferentes categorias deve ser igual ao total de ensaios. O termo  $\binom{N}{\vec{n}}$  é o coeficiente multinomial, uma generalização do coeficiente binomial para ensaios com mais de dois resultados possíveis:  $\frac{N!}{n_1!n_2!\dots n_k!}$ . A distribuição multinomial atribui probabilidade a um vetor com  $K$  dimensões,  $\vec{n}$ , no qual cada elemento  $n^{(k)}$  contabiliza quantas vezes um evento  $k$  aconteceu num total de  $N$  ensaios. O vetor  $\vec{p}$  também é composto por  $K$  dimensões e indica a probabilidade de um evento  $k$  acontecer num ensaio isolado.

Quando a distribuição multinomial é reduzida a um único ensaio, ela costuma ser simplificada e passa a ser denominada *distribuição categorial* ou *discreta*. Nessa distribuição, o vetor  $\vec{n}$  é substituído pelo índice correspondente ao evento  $k$ ,  $z$ ; e o número de ensaios desaparece.

$$p(z \mid \vec{p}) = p_z \triangleq \text{Cat}(z \mid \vec{p}) \quad (2.14)$$

A distribuição *a priori* conjugada associada ao parâmetro  $\vec{p}$  é a *distribuição de Dirichlet*, uma generalização da distribuição Beta de duas para  $K$  dimensões (Gelman et al., 2014). Sua parametrização é dada pelo parâmetro de concentração  $\vec{\alpha}$ , que pode ser interpretado no caso do uso conjugado com a distribuição multinomial como a contagem de ocorrência de cada resultado *a priori*, ou pseudocontagens. Sua formulação é, basicamente, uma generalização da distribuição multinomial para aceitar números reais, substituindo o uso de fatoriais pela função gama ( $\Gamma$ ). Muitas vezes o parâmetro  $\vec{\alpha}$  é

composto por elementos de igual valor. Nesse caso, essa distribuição é denominada de *distribuição de Dirichlet simétrica*.

$$p(\vec{p} \mid \vec{\alpha}) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K p_k^{\alpha_k - 1} \triangleq \text{Dir}(\vec{p} \mid \vec{\alpha}) \quad (2.15)$$

A distribuição de Dirichlet especifica vetores num *simplex* de dimensionalidade  $K - 1$ , forçando com que a soma dos elementos do vetor representado seja sempre igual a 1. Quando a distribuição de Dirichlet é utilizada como distribuição *a priori* em conjunto com uma função de verossimilhança multinomial ou categorial, a distribuição posterior do parâmetro  $\vec{p}$  tem a mesma representação da distribuição de Dirichlet utilizada *a priori*, com a única diferença que a distribuição posterior soma as pseudocontagens *a priori* com a contagem de ocorrências observadas de cada categoria.

$$\begin{aligned} p(\vec{p} \mid \vec{n}, \vec{\alpha}, N) &\propto p(\vec{p} \mid \vec{\alpha}, N) p(\vec{n} \mid \vec{p}, \vec{\alpha}, N) \\ &= \text{Dir}(\vec{p} \mid \vec{\alpha}) \text{Mult}(\vec{n} \mid \vec{p}, N) \\ &= \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K p_k^{\alpha_k - 1} \binom{N}{\vec{n}} \prod_{k=1}^K p_k^{n^{(k)}} \\ &= \frac{\Gamma(\sum_{k=1}^K \alpha_k + n^{(k)})}{\prod_{k=1}^K \Gamma(\alpha_k + n^{(k)})} \prod_{k=1}^K p_k^{\alpha_k + n^{(k)} - 1} \\ &= \text{Dir}(\vec{p} \mid \vec{\alpha} + \vec{n}) \end{aligned} \quad (2.16)$$

Na equação acima, a distribuição posterior, do lado esquerdo da equação, é primeiramente definida em termos da distribuição *a priori*, identificada como uma distribuição de Dirichlet com hiperparâmetro  $\vec{\alpha}$ , e da verossimilhança, identificada como uma distribuição multinomial. Como a distribuição de Dirichlet é conjugada da distribuição multinomial, a terceira linha pode ser facilmente resolvida desconsiderando as constantes dadas pela função Beta (primeiro termo da distribuição de Dirichlet) e pelo termo multinomial, somando os expoentes do vetor  $\vec{p}$  tanto na distribuição *a priori* quanto na função de verossimilhança, e renormalizando com a função beta atualizada, como na quarta linha.

Apesar do exemplo utilizado aqui ser um pouco mais complicado do que o necessário para demonstrar o funcionamento da estimação bayesiana por meio de distribuições conjugadas, ele se



mostrará útil mais tarde quando abordarmos os modelos de tópicos. O modelo de tópico mais importante, a Alocação de Dirichlet Latente, faz uso das distribuições multinomial e Dirichlet, bem como se aproveita de sua conjugabilidade, para modelar o processo de geração de dados textuais em documentos.

#### 2.1.4.2 Monte Carlo por Cadeias de Markov

A conjugabilidade, porém, apresenta limites importantes. Na maioria dos problemas aplicados, a decisão de utilizar uma distribuição *a priori* conjugada da função de verossimilhança se deve à expectativa de facilitar o procedimento de inferência. Infelizmente, isso nem sempre é aceitável, quando o sistema modelado não é bem representado pela função de verossimilhança escolhida, ou quando nosso conhecimento prévio sobre um parâmetro não pode ser expresso numa distribuição paramétrica conjugada. Nesses casos, a solução é utilizar funções de verossimilhança e distribuições *a priori* não conjugadas, o que torna necessário calcular o denominador da regra de Bayes. Para lidar com a intratabilidade do cálculo da evidência nesses casos, é necessário utilizar métodos aproximativos. Um dos métodos mais utilizados para aproximar a distribuição posterior de um modelo Bayesiano é a simulação de Monte Carlo por Cadeias de Markov (*Markov Chain Monte Carlo*, MCMC) (Gelman et al., 2014).

A principal ideia por trás do MCMC é substituir a distribuição posterior por amostras retiradas dessa distribuição. Na medida em que essas amostras tendem ao infinito, as estatísticas calculadas a partir da amostra coincidem assintoticamente com os parâmetros de interesse da distribuição posterior, aspecto garantido pela Lei dos Grandes Números. A principal vantagem do uso de MCMC para simular valores de distribuições de probabilidade oriundas da utilização da regra de Bayes é que não é necessário computar o denominador ou a evidência. Isso porque as simulações automaticamente geram valores proporcionalmente à sua probabilidade, permitindo a normalização da distribuição (Gelman et al., 2014; Kruschke, 2010).

Mas o que é exatamente uma cadeia de Markov? Uma sequência de variáveis aleatórias na qual cada variável é independente das demais quando condicionada à variável anterior na sequência.

$$p(\mathcal{X}_i \mid \mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_{i-1}) = p(\mathcal{X}_i \mid \mathcal{X}_{i-1}) \quad (2.17)$$

A distribuição de probabilidade condicional de numa cadeia

de Markov é definida pelo *núcleo de transição* da cadeia, uma função ou matriz que estabelece com qual probabilidade ela pode passar para outro estado  $\mathcal{X}_i$  (ou permanecer no mesmo estado), dado que a cadeia se encontra num estado  $\mathcal{X}_{i-1}$ . Quando a cadeia de Markov é ergódica, ou seja, se ela pode passar de um estado para qualquer outro, independentemente do estado atual, ela converge para uma distribuição estacionária quando simulada um número suficientemente grande de vezes. Graças a essa propriedade, cadeias de Markov podem ser utilizadas para se chegar a uma aproximação de qualquer distribuição de probabilidade, inclusive distribuições posteriores (Kruschke, 2010).

A possibilidade da cadeia de Markov chegar a essa distribuição estacionária, porém, depende do núcleo de transição escolhido para a cadeia. Para que a distribuição estacionária seja a distribuição de probabilidade buscada pelo processo de inferência, a probabilidade da cadeia se encontrar num determinado estado deve ser invariante com relação ao núcleo de transição escolhido (Griffiths et al., 2008). Um dos algoritmos mais conhecidos para a criação de cadeias de Markov com a propriedade de convergir para a distribuição de probabilidade de interesse é o *Metropolis*. Nesse algoritmo, valores novos dos parâmetros são propostos a partir de uma distribuição simétrica ( $\theta^*$ ), condicionada ao valor dos parâmetros do momento anterior na cadeia ( $\theta^{t-1}$ ). Em seguida, é calculada uma probabilidade de aceitação, definida por  $\min(1, \frac{p(\theta^*)}{p(\theta^{t-1})})$ , que por sua vez é comparada com um sorteio aleatório de uma distribuição entre 0 e 1. Se a probabilidade de aceitação é maior que o sorteio aleatório, a cadeia muda seu estado para o valor proposto; caso contrário, permanece no valor do estado atual. Como fica óbvio pela definição da probabilidade de aceitação, a cadeia sempre muda de estado quando os novos parâmetros propostos tem maior probabilidade, mas pode também retornar a regiões de menor probabilidade, o que garante a convergência assintótica entre a distribuição posterior e a distribuição estacionária da cadeia de Markov (Gelman et al., 2014; Kruschke, 2010).

Outra técnica bastante utilizada para gerar cadeias de Markov é a *amostragem de Gibbs*. Nesse algoritmo, não há valores propostos e probabilidade de aceitação; em seu lugar, os valores de uma variável são amostradas a partir da distribuição dessa variável condicionada ao valor atual de todas as demais variáveis (Gelman & Hill, 2006). A amostragem de Gibbs é particularmente útil no caso de utilização de distribuições conjugadas, pois facilitam a determi-

nação das distribuições condicionais. Além disso, não é necessário condicionar os valores de uma variável a todas as outras, mas apenas naquelas variáveis que fazem parte do *cobertor de Markov* da variável amostrada. Essa noção será abordada logo abaixo, por ser mais fácil de ser compreendida em termos dos modelos gráficos probabilísticos.

#### 2.1.4.3 Inferência variacional

Uma das grandes desvantagens da utilização de simulações para aproximar a distribuição posterior é que, apesar de garantias teóricas de convergência assintótica para a distribuição alvo, na prática, as cadeias de Markov podem ficar presas e não conseguir explorar suficientemente toda a distribuição de interesse. Além disso, os métodos de simulação exigem grande potência computacional e não são indicados para grande volume de dados ou modelos excessivamente complicados, especialmente modelos com variáveis latentes, que costumam ter a distribuição posterior multimodal (Gelman et al., 2014).

No contexto dos modelos de tópicos e outros modelos de mistura – exemplos de modelos com variáveis latentes – uma alternativa eficiente para trabalhar com distribuições complexas e grande quantidade de dados é a *inferência variacional*. Em vez de simular dados da distribuição posterior de interesse, geralmente de difícil tratamento analítico, a inferência variacional propõe uma família de distribuições de probabilidade com estrutura mais simples e tratável, formulando o problema de estimar a distribuição posterior como uma tarefa de otimização: encontrar, dentre a família de distribuições mais simples proposta, qual distribuição minimiza a perda de informação com relação à distribuição posterior de interesse (Blei et al., 2003).

Em outras palavras, a inferência variacional tenta firmar um compromisso entre a tratabilidade da distribuição posterior e a precisão de seu ajuste. Por trabalhar com uma família de distribuições mais simples que a distribuição alvo, o resultado da inferência será sempre aproximado – diferentemente do uso de MCMC, que pode aproximar a distribuição posterior exata até um nível arbitrariamente pequeno de erro. Em compensação, o ajuste de um modelo por inferência variacional tende a ser bem mais rápido, chegando a ordens de magnitude de diferença, o que permite utilizá-lo em situações em que a utilização de métodos de simulação seria impraticável

(Bishop, 2006).

Para implementar a inferência variacional, são necessários três elementos: um *funcional*, ou seja, uma função que tem outras funções como argumento e retorna um escalar; uma representação formal da distribuição posterior que será aproximada; e, por fim, uma família de distribuições aproximadas da distribuição posterior, denominadas de *distribuições variacionais*. O principal funcional utilizado na literatura sobre inferência variacional é a divergência de Kullback-Leibler, uma medida que foi desenvolvida na teoria da informação e que mensura a quantidade de informação que é perdida quando se utiliza uma distribuição de probabilidade  $q$  para aproximar uma distribuição  $p$ . No caso de distribuições de probabilidade discretas, a definição é dada pela Equação 2.18. A mesma definição é válida para funções de densidade de probabilidade, trocando a soma por uma integral (Bishop, 2006).

$$D_{\text{KL}}(p||q) = \sum_i p(i) \log \frac{p(i)}{q(i)} \quad (2.18)$$

A divergência de Kullback-Leibler é minimizada quando  $p = q$ , e, portanto, minimizar a distância entre duas distribuições significa encontrar, dentro dos limites estabelecidos para a distribuição variacional aproximativa, qual distribuição melhor representa a distribuição alvo. Em muitos casos de inferência variacional bayesiana, a divergência de Kullback-Leibler não pode ser otimizada diretamente pela presença de integrais intratáveis no denominador da distribuição posterior. Nesses casos, é usual utilizar uma divergência derivada denominada de *limite inferior da evidência* (*evidence lower bound*, ELBO), que é proporcional ao negativo da divergência de Kullback-Leibler e leva ao mesmo resultado de otimização quando maximizada (Blei, Kucukelbir, & McAuliffe, 2016).

A distribuição a ser aproximada é a distribuição posterior dos parâmetros do modelo condicionada aos dados observados, tal como descrevemos em linhas gerais acima quando falamos da inferência bayesiana. A família de distribuição variacionais, por sua vez, costuma ser baseada na fatoração da distribuição posterior de maneira a representar cada parâmetro por uma distribuição independente. Essa aproximação, tomada de empréstimo da Física, onde recebe o nome de *Teoria do Campo Médio*, divide os parâmetros do modelo em  $K$  blocos independentes entre si (Bishop, 2006):

$$q(\vec{\theta}) = \prod_{i=1}^K q_i(\theta_i) \quad (2.19)$$

Ou seja, para tornar o cálculo mais tratável, a distribuição aproximativa  $q(\vec{\theta})$  fatora os parâmetros em conjuntos que podem ser avaliados independentemente uns dos outros. O grau de complexidade da distribuição variacional é arbitrário, mas precisa levar em conta a precisão desejada para a aproximação. Quanto mais complexa a distribuição variacional, maior a possibilidade de aproximar a distribuição posterior com pouco erro, mas maior a dificuldade em minimizar a divergência entre as duas distribuições. É importante ressaltar que a forma paramétrica de cada  $q_i(\theta_i)$  não é definida de antemão, mas é encontrada pelo processo de otimização de acordo com a estrutura da distribuição conjunta a ser aproximada (Blei et al., 2016).

Para a minimização da divergência, ou, de maneira equivalente, a maximização do limite inferior da evidência, o algoritmo mais comumente utilizado é *ascensão de coordenadas* (*coordinate ascent*). Neste algoritmo, cada bloco de parâmetros é inicializado aleatoriamente ou por algum critério aproximativo e atualizado iterativamente assumindo os outros blocos como fixos em seus valores atuais. Ou seja, a cada iteração, os blocos de parâmetros são atualizados utilizando os valores computados para os outros blocos na iteração anterior. Essa estratégia tem conversão garantida, mas apenas para um valor ótimo local (Bishop, 2006; Blei et al., 2016).

## 2.2 Modelos gráficos

A abordagem feita até agora da inferência bayesiana não deixou claro como especificar o modelo cujos parâmetros serão estimados a partir dos dados. Comentamos como a regra de Bayes permite calcular os valores mais prováveis de um conjunto de parâmetros a partir de dados observados. Mas como definir que parâmetros são necessários para modelar um fenômeno?

A forma mais simples, muitas vezes utilizada para fenômenos pouco complexos, é simplesmente escolher uma distribuição de parametrização conhecida e utilizar a regra de Bayes para calcular a distribuição posterior dos parâmetros. A maioria dos fenômenos que estudamos em Psicologia, todavia, não são facilmente descritos por apenas uma distribuição isolada. Pelo contrário, os modelos

mais verossímeis tendem a ser muito mais complexos, compostos por um número arbitrário de variáveis com relações hierárquicas específicas entre si. Esses modelos não são facilmente reduzidos a uma distribuição conhecida, mas precisam ser definidos como uma distribuição conjunta de variáveis observadas e latentes. Quando o número de variáveis torna difícil a compreensão de um modelo, uma alternativa é utilizar modelos gráficos probabilísticos. Com o auxílio de modelos gráficos é possível construir um modelo a partir de módulos que se repetem e relações de dependência claramente especificadas (Griffiths et al., 2008), como é o caso dos modelos de tópicos.

Num modelo gráfico, a distribuição de probabilidade do modelo é associada a um grafo. Os vértices presentes no grafo representam variáveis, que podem ser observadas ou latentes – no referencial bayesiano, mesmo os parâmetros de um modelo são tratados como variáveis aleatórias não observadas. As arestas que ligam os vértices indicam relações de dependência entre variáveis, o que torna o modelo gráfico um mapa das dependências das variáveis e permite identificar também relações de independência. A dependência entre variáveis indicada pelas arestas é acompanhada de uma função que torna explícita sua relação – na maioria das vezes, uma função de densidade de probabilidade paramétrica conhecida (Heinrich, 2009).

Os modelos gráficos costumam ser classificados em dois tipos principais: aqueles cujas arestas não são direcionadas, indicando apenas a dependência mútua entre variáveis – denominados de *modelos gráficos não-direcionados*, ou de *Campos aleatórios de Markov* (*Markov Random Fields*); e aqueles cujas arestas são direcionadas, indicando uma hierarquia entre os vértices do modelo – chamados de *modelos gráficos direcionados*, ou *redes bayesianas* (*Bayesian Networks*).

### 2.2.1 Redes bayesianas

Uma rede bayesiana, como qualquer outro modelo gráfico probabilístico, representa as dependências existentes entre um conjunto de variáveis. Como são compostas por arestas direcionadas, o vértice de onde parte a aresta é denominado de “vértice pai” e o vértice sobre o qual incide a aresta recebe o nome de “vértice filho”. Mantendo a nomenclatura genealógica, são utilizados termos como “vértices ancestrais” e “vértices descendentes” quando uma rede é composta por vários níveis hierárquicos. Para garantir a identifica-

bilidade da distribuição representada pela rede bayesiana, a relação entre as variáveis precisa ser acíclica, ou seja, não é possível retornar a um mesmo vértice seguindo a direção das arestas. Por esse motivo, as redes bayesianas também são chamadas de *Grafos Direcionados Acíclicos* (*Directed Acyclic Graph*, DAG). No caso de sistemas mais complexos composto por múltiplas variáveis que se repetem, é usual representar a repetição por meio de travessas ou pratos (*plates*), tornando a representação visual ainda mais enxuta que a definição formal do modelo (Griffiths et al., 2008). Ilustraremos a utilização de travessas quando apresentarmos a Alocação de Dirichlet Latente e alguma de suas alternativas.

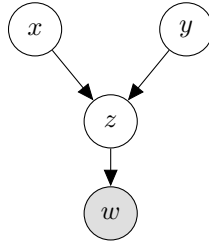
Em suma, as redes bayesianas constituem uma linguagem gráfica formal que permite representar a distribuição conjunta de um sistema em termos de variáveis aleatórias, bem como suas dependências, num grafo cujas arestas são todas direcionadas e não formam ciclos. Graças à representação das principais dependências da distribuição conjunta, as redes bayesianas simplificam consideravelmente os cálculos de probabilidade (Heinrich, 2009).

A simplificação do cálculo da distribuição conjunta das variáveis de uma rede bayesiana deve-se à sua consistência com a *condição de Markov*. Isso quer dizer que, quando condicionada a seus vértices pais, todas as variáveis são independentes umas das outras, com exceção de suas variáveis descendentes. Em vez de aplicar a regra da cadeia em toda sua extensão (ou seja, condicionando cada variável a todas as outras variáveis anteriormente consideradas), a rede bayesiana permite fatorar a distribuição de probabilidade conjunta como o produto de distribuições localizadas em que cada variável é condicionada apenas às suas variáveis pais (Griffiths et al., 2008).

Para exemplificar melhor o que isso significa em termos da redução da complexidade de uma distribuição conjunta, considere o conjunto de quatro variáveis aleatórias  $\{X, Y, Z, W\}$ . Pela regra da cadeia, desconsiderando qualquer estrutura de relações entre essas variáveis, a distribuição conjunta poderia ser definida da seguinte maneira:

$$p(x, y, z, w) = p(x)p(y | x)p(z | x, y)p(w | x, y, z) \quad (2.20)$$

Suponhamos, porém, que a distribuição dessas quatro variáveis possa ser representada por uma rede bayesiana, conforme a figura abaixo.



**Figura 2.1:** Exemplo de modelo gráfico probabilístico

Graças à condição de Markov, a distribuição conjunta agora pode ser representada de maneira muito mais econômica:

$$p(x, y, z, w) = p(x)p(y)p(z \mid x, y)p(w \mid z) \quad (2.21)$$

A estrutura da rede explicita que, na distribuição conjunta das variáveis, as variáveis mais ancestrais,  $X$  e  $Y$ , são independentes de todas as outras variáveis e, portanto, não precisam ser condicionadas. A variável  $Z$  depende tanto de  $X$  quanto de  $Y$ , indicando a necessidade de condicionamento. Por fim, a variável  $W$  depende diretamente apenas de  $Z$ , e só precisa ser condicionada a ela. A vantagem dessa representação pode não ficar evidente para um modelo simples, mas torna-se muito mais prática quando o número de variáveis aumenta, como ficará evidente nos modelos de tópicos.

### 2.2.2 Dependência, independência e independência condicional

Um dos grandes méritos das redes bayesianas, portanto, é sua habilidade em codificar as relações de dependência entre variáveis. Com isso, é possível inferir também as relações de independência, e, em particular, as independências condicionais. Duas variáveis são consideradas independentes se informações sobre uma delas não permite inferir novas informações sobre a outra. Isso pode ser representado pela regra da cadeia para variáveis independentes:  $p(X, Y) = p(X)p(Y)$  ou então  $X \perp\!\!\!\perp Y$ . As redes bayesianas permitem inferir relação de independência condicional, ou seja, duas variáveis que são marginalmente dependentes tornam-se independentes quando condicionadas a uma terceira variável. Sua notação implica



no condicionamento das variáveis:  $p(X, Y \mid Z) = p(X \mid Z)p(Y \mid Z)$ , ou então  $X \perp\!\!\!\perp Y \mid Z$  (Heinrich, 2009).

Há duas maneiras gerais para se determinar a independência condicional de uma variável. A primeira é denominada de *cobertor de Markov* (*Markov blanket*). Essa regra especifica que uma variável é independente de todas as outras variáveis da rede quando condicionada ao seu cobertor de Markov. O cobertor de Markov, por sua vez, é um subgrafo da rede completa composta pelos vértices pais, filhos e co-pais (outros pais dos vértices filhos) de um determinado vértice. Portanto, se definimos o cobertor de Markov de uma variável  $X_i$  como  $B(X_i)$ , podemos determinar que  $X_i \perp\!\!\!\perp X_{-i} \mid B(X_i)$ , ou seja, que a variável  $X_i$  é independente de todas as outras variáveis do grafo,  $X_{-i}$ , quando condicionada ao seu cobertor de Markov (Heinrich, 2009).

A segunda maneira para determinar a independência condicional entre variáveis é pelo conjunto de vértices não-descendentes de uma variável. Se ordenarmos os vértices do grafo de tal forma que nenhum vértice apareça antes de seus pais – a *ordem topológica* do grafo – todos os vértices que aparecem antes de um determinado vértice e que não são seus pais são denominados de *não-descendentes*. Podemos determinar que a variável  $X_i$  é independente de seus não-descendentes  $D(X_i)$  quando condicionado a seus pais  $P(X_i)$ . Com isso, obtemos  $X_i \perp\!\!\!\perp D(X_i) \mid P(X_i)$  (Heinrich, 2009).

Uma maneira mais intuitiva de avaliar a independência condicional entre vértices de uma rede bayesiana é por meio do método da *bola bayesiana* (*Bayes ball*) (Shachter, 2013). Esse método verifica a possibilidade de passar uma mensagem qualquer entre os vértices de um grafo seguindo um conjunto de regras simples que leva em consideração apenas relações imediatas entre variáveis, mas que pode ser concatenado para verificar a dependência em grafos mais complexos.

Quando estamos lidando com um grafo ou um subgrafo em que nenhuma informação é observada, ou seja, não há um nodo condicionante, duas variáveis  $X$  e  $Y$  sempre serão dependentes se: estiverem diretamente ligadas (uma é pai da outra,  $X \rightarrow Y$  ou  $Y \rightarrow X$ ); estiverem conectadas por meio de uma terceira variável  $Z$ , também não observada, numa cadeia ( $X \rightarrow Z \rightarrow Y$ ); ou estiverem conectadas por meio de uma terceira variável que é pai de ambas (formando um estrutura em  $A$ ,  $X \leftarrow Z \rightarrow Y$ ). O único caso em que as duas variáveis serão independentes é no caso de estarem ligadas por uma mesma variável que é filha de ambas (formando

uma estrutura em  $V$ ,  $X \rightarrow Z \leftarrow Y$ ) (Shachter, 2013).

Quando estamos considerando o caso de um subgrafo ou grafo em que há pelo menos um vértice observado, as regras de dependência mudam de acordo com a posição do vértice condicionante. No caso da ligação direta de duas variáveis, seja uma delas observada ou não, sempre serão dependentes. Por outro lado, se a variável observada  $Z$  estiver entre  $X$  e  $Y$  numa organização de cadeia,  $X$  e  $Y$  serão independentes *quando condicionadas em  $Z$* . Da mesma forma, se a variável de ligação  $Z$  for observada e for pai de  $X$  e  $Y$ , formando uma estrutura em  $A$ ,  $X$  e  $Y$  também serão independentes quando condicionadas em  $Z$ . O último caso, quando a variável de ligação  $Z$  for observada e filha de  $X$  e  $Y$ , numa estrutura em  $V$ , as variáveis se tornam dependentes quando condicionadas em  $Z$  (Shachter, 2013).

Em outras palavras, “as regras da bola de Bayes especifica que vértices filhos bloqueiam a propagação se e somente se eles são latentes, enquanto vértices pais e transicionais bloqueiam a propagação se e somente se eles estão dados ou são observados” (Heinrich, 2009, p. 14). Se não há nenhum caminho ativo entre duas variáveis, condicionadas às variáveis observadas ao longo do caminho, elas são independentes e são chamadas de *d-separadas* pelas variáveis condicionantes (Bishop, 2006).

### 2.2.3 Modelos gráficos como modelos gerativos

Uma rede bayesiana permite formalizar o processo pelo qual os dados observados foram gerados. Seguindo a ordem hierárquica explicitada pelo modelo, cada variável é amostrada a partir de sua distribuição condicionada aos valores de suas variáveis pais. Por exemplo, no modelo hipotético da Figura 2.1, primeiramente valores para as variáveis  $X$  e  $Y$  são sorteadas das distribuições  $p(X)$  e  $p(Y)$ . Em seguida, respeitando a estrutura hierárquica, um valor para a variável  $Z$  é sorteado da distribuição condicional  $p(Z | X, Y)$ . Por fim, um valor para  $W$  é sorteado da distribuição condicional  $p(W | Z)$ . Essa interpretação de uma rede bayesiana é denominada de *modelo gerativo*, pois indica as etapas necessárias para a geração dos valores de um sistema de variáveis (Griffiths et al., 2008).

Para fundamentar o uso de redes bayesianas como modelos gerativos, é necessária uma noção de independência mais forte que a independência condicional, a *intercambialidade* (*exchangeability*). Esse pressuposto assume que uma sequência finita de variáveis ale-

atórias é intercambiável se sua distribuição conjunta é invariante a qualquer permutação de sua ordem. Numa sequência infinita de variáveis aleatórias, a intercambialidade requer que qualquer subsequência finita seja intercambiável.

A importância dessa noção se deve à garantia dada por ela que, uma vez condicionada a um parâmetro sorteado de uma distribuição *a priori*, uma sequência de variáveis aleatórias pode ser modelada como condicionalmente independente. Esse resultado da teoria da probabilidade é conhecido como *Teorema de Representação de de Finetti* e costuma ser utilizado como justificativa matematicamente fundada para a utilização de distribuições *a priori* na inferência bayesiana. O princípio de intercambialidade também justifica o uso da notação de travessa nos modelos gráficos, pois, uma vez sorteado um parâmetro que determina a distribuição das variáveis sobre as travessas, os valores sorteados para essas variáveis podem ser considerados independentes quando condicionados ao parâmetro do vértice pai (Heinrich, 2009). A Alocação de Dirichlet Latente, que abordaremos em maiores detalhes a seguir, faz uso do princípio de intercambialidade e do teorema de representação, ao assumir que as palavras presentes em um documento são infinitamente intercambiáveis (Blei et al., 2003).

Além dessas propriedades teóricas que justificam o uso de redes bayesianas para representar a distribuição conjunta de um grupo de variáveis, ela também permite representar de maneira intuitiva o processo teórico concebido como sendo aquele que gerou os dados observados. As relações entre variáveis pode inclusive representar o fluxo de causalidade entre variáveis, quando isso for justificado para o fenômeno estudado.

Utilizando a estrutura codificada na rede bayesiana sobre o processo que levou à geração dos dados, a inferência bayesiana permite inverter o processo e inferir, condicionando nos dados observados, os valores mais prováveis dos parâmetros e variáveis latentes (Blei, 2012). Em função disso, é possível ajustar um modelo teórico a um conjunto de dados observados. Essa estratégia será utilizada na análise de dados textuais a partir dos modelos de tópicos: proporemos uma estrutura teórica sobre como relações semânticas entre as palavras determinam a formação de documentos, e aplicaremos a inferência bayesiana, calculando a distribuição posterior dos parâmetros latentes por métodos aproximativos como MCMC, para estimar os parâmetros dos modelos a partir de um *corpus*.

## 2.3 Alocação de Dirichlet Latente

Tendo esclarecido alguns conceitos fundamentais necessários à compreensão dos modelos de tópicos, passamos a abordar o trabalho seminal de David Blei (Blei et al., 2003), base que utilizaremos para modelar o processo de associação livre. Apesar de não poder ser considerado a primeira formulação de um modelo de tópico, o trabalho de Blei serviu de referência central para todas as investigações posteriores na área. Mas para compreender a origem do modelo, é necessário situar quais problemas ele buscou resolver e quais trabalhos o antecederam.

### 2.3.1 Problema motivador e antecedentes

O aumento significativo no acervo digital do conhecimento humano torna cada vez mais trabalhoso encontrar informações de interesse. O principal motivo dessa dificuldade é a ausência de uma estrutura bem organizada na maior parte dos dados digitalizados, o que impossibilita uma categorização clara e, por consequência, dificulta sua recuperação em meio a outras informações. Apesar de esses documentos pouco estruturados poderem assumir variadas formas – vídeos, áudios, imagens – a maioria do conhecimento digitalizado é constituído por dados textuais. Para lidar com a pouca estrutura dos textos, os mecanismos de busca e descoberta de informação identificam a presença ou ausência de termos de pesquisa no conteúdo dos documentos. Esses termos costumam se limitar à correspondência de palavras exatas utilizadas em cada busca. O interesse em investigar um conjunto de documentos, porém, não se restringe à busca de palavras específicas, mas também é motivado por significados e temas a elas associados (Blei & Lafferty, 2009).

A área de *Information Retrieval*, ramo das Ciências da Computação que investiga métodos de recuperação de informações relevantes a partir de diversas fontes de dados, desenvolveu várias estratégias para lidar com a limitação da busca e exploração de conjuntos pouco estruturados de documentos (Turney & Pantel, 2010). Uma estratégia para a busca por documentos textuais se dá pela redução de cada documento de uma coleção a um vetor de números reais. Esse vetor é extraído de uma matriz que apresenta as contagens de cada termo nos documentos de uma coleção, chamada de *matriz de termos-documentos*. Nessa matriz, cada linha representa um termo do vocabulário, cada coluna representa um documento da

coleção, e cada célula apresenta a contagem de ocorrências de um termo dentro de um documento. A contagem bruta da frequência dos termos, porém, costuma ser enviesada, pois palavras de pouco valor semântico, como verbos auxiliares, conjunções e preposições costumam ter uma incidência muito maior do que palavras significativas, como substantivos e adjetivos. Esse problema geralmente é endereçado por uma transformação da contagem da frequência dos termos presentes nos documentos. Uma das principais transformações utilizadas é a *tf-idf* (*term frequency - inverse document frequency*), que leva em consideração tanto o número de vezes que um termo aparece no documento quanto a representatividade daquele termo para o documento. A intuição por trás dessa estratégia é simples: um documento no qual um termo aparece várias vezes deverá ser mais pertinente à busca; por outro lado, um termo que aparece com muita frequência em vários documentos provavelmente é pouco informativo da especificidade de cada documento (Blei et al., 2003).

Um avanço da utilização da matriz bruta produzida pela aplicação da *tf-idf* é a utilização de técnicas de redução de dimensões, como a decomposição de valores singulares (SVD), para identificar representações de menor dimensionalidade da matriz de termos e documentos. Essa técnica ficou conhecida como *Indexação Semântica Latente* (*Latent Semantic Indexing*, LSI) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990), e permite reduzir a matriz de termos e documentos a um conjunto de matrizes que capturam a maior parte da variação. Além de permitir uma grande compressão da matriz original, a técnica também captura aspectos semânticos como sinonímia e polissemia, pois palavras que aparecem em contextos semelhantes tendem a se agrupar na representação de menor dimensionalidade. O resultado alcançado pela LSI na desambiguação de termos homônimos e na aproximação de palavras semanticamente próximas levou a sua aplicação como modelo para a memória semântica e o processamento significativo da linguagem (Landauer, Foltz, & Laham, 1998), passando a ser denominada na Psicologia Cognitiva como *Análise Semântica Latente*.

Como estratégia para avaliar o desempenho da LSI, foi desenvolvido um modelo gerativo probabilístico para documentos textuais (Papadimitriou, Raghavan, Tamaki, & Vempala, 1997). O objetivo desse primeiro modelo é aplicar a probabilidade como uma estratégia de avaliação dos resultados obtidos pela aplicação da SVD sobre matrizes de termos e documentos. Essa iniciativa, porém,

acabou levando ao desenvolvimento de modelos probabilísticos ajustados diretamente a partir dos dados. Um dos principais modelos desenvolvidos a partir da LSI foi o *probabilistic Latent Semantic Indexing* (pLSI) (Hoffmann, 1999), também conhecido como *aspect model*. Na pLSI, cada palavra de um documento provém de um modelo de mistura composto por uma distribuição de proporção de tópicos. Ou seja, cada palavra provém de um tópico e o documento é formado pela mistura de vários tópicos. Com isso, os documentos podem ser representados de forma reduzida pela distribuição da proporção de tópicos. Esse modelo, porém, apresenta limitações: ele não apresenta um modelo gerativo completo para as proporções de tópicos num documento. A pLSI foi a primeira estratégia de recuperação de informação desenvolvida que usa modelagem probabilística para identificar um conjunto de temas latentes num *corpus* textual, e pode ser considerada como o primeiro modelo de tópicos.

A principal vantagem dos modelos de tópicos sobre outras estratégias de redução de dados textuais é permitir explorar a estrutura latente de temas que constituem um conjunto de documentos compostos por textos, inclusive organizando uma estrutura hierárquica entre tópicos mais amplos e mais específicos. Essa organização dos tópicos e suas proporções nos documentos também possibilita identificar as mudanças ocorridas ao longo do tempo e a relação dos tópicos entre si. Encontrando um tema de interesse, é possível investigar o conjunto de documentos relacionados a ele. Isso torna possível identificar e classificar automaticamente grandes conjuntos de documentos com informações sobre seus temas, não havendo necessidade de que os documentos tenham anotações ou rótulos descritivos prévios, pois os temas são identificados a partir do próprio conteúdo dos textos (Blei, 2012).

Abordaremos em seguida a *Alocação de Dirichlet Latente* (*Latent Dirichlet Allocation*, LDA), modelo seminal que provocou um grande aumento nas pesquisas sobre modelos de tópicos. O modelo será descrito em suas definições como *modelo probabilístico gerativo*, como *distribuição posterior de probabilidade* e como *modelo gráfico probabilístico*.

### 2.3.2 Notação e terminologia

Para descrever as variáveis envolvidas na LDA, utilizaremos a notação empregada por Blei e Lafferty (2009) em conjunto com as definições do artigo original sobre a LDA (Blei et al., 2003).

1. As seguintes variáveis são diretamente observadas:

- Um *termo* é uma unidade básica dos dados discretos, definida como um item proveniente do vocabulário indexado por  $\{1, \dots, V\}$ . Um termo é representado como um vetor com dimensionalidade  $V$  no qual todos os elementos são zero com exceção do  $v$ -ésimo elemento, que é um. Ou seja, um termo é um vetor  $w$  tal que  $w^v = 1$  e  $w^u = 0$ , para todo  $u \neq v$ .
  - Utilizaremos o conceito *termo* para se referir a cada palavra única utilizada nos documentos de uma coleção e que compõe o vocabulário;
  - O vocábulo *palavra* será utilizado para se referir às instâncias de um termo nos documentos do *corpus*.
- Um *documento* é uma sequência de  $N$  palavras denotado por  $\vec{w} = (w_1, w_2, \dots, w_N)$ , de tal forma que  $w_n$  é a  $n$ -ésima palavra na sequência e cada palavra é uma instância de um termo do vocabulário  $V$ .
- Um *corpus* é uma coleção de  $D$  documentos representado por  $\mathcal{D} = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_D\}$ .

2. Os seguintes parâmetros são latentes e inferidos pelo modelo:

- Um *tópico* é uma distribuição multinomial sobre o vocabulário e é representado como  $\vec{\beta}$ . O modelo é composto por um número fixo de tópicos assumido *a priori*,  $K$ , e o conjunto completo de tópicos é representado por  $\vec{\beta}_{1:K}$ .
- Cada palavra  $w_n$  de um documento é considerada como pertencendo a um dos tópicos do modelo, e a atribuição de um tópico para uma palavra é representada por  $z_{n,d}$ , ou seja, a  $n$ -ésima palavra do documento  $d$  foi retirada do tópico  $z$ , um dos  $K$  tópicos do modelo.
- Cada documento é concebido como uma mistura de diferentes proporções de todos os tópicos da coleção, e a proporção de mistura do documento  $d$  é indicada pelo vetor com  $K$  dimensões  $\vec{\theta}_d$ .

3. Os seguintes hiperparâmetros são fixos *a priori* ou estimados, de acordo com o algoritmo utilizado:

- A distribuição *a priori* dos tópicos da coleção é dada por uma distribuição de Dirichlet simétrica de dimensão  $V$  e

com parâmetro de concentração  $\eta$  – sendo uma distribuição simétrica, a concentração de cada dimensão tem igual valor.

- A distribuição *a priori* da proporção de tópicos para cada documento também segue uma distribuição de Dirichlet (que pode ser simétrica ou não) com dimensão  $K$  e parâmetro de concentração  $\vec{\alpha}$ .

### 2.3.3 A LDA como modelo gerativo

A LDA é um modelo relativamente simples que compreende os documentos como constituídos por uma série de tópicos diversos (Blei et al., 2003). Mais especificamente, a LDA assume que cada documento num *corpus* é gerado por uma mistura de diferentes proporções de um número limitado de tópicos, e cada tópico contribui com várias palavras associadas a ele.

No contexto da LDA, um *tópico* é definido como uma distribuição multinomial<sup>3</sup> sobre um vocabulário estático, retirada de uma distribuição de Dirichlet informada *a priori*. Ou seja, um tópico é constituído por uma série fixa de palavras às quais são associadas probabilidades de acordo com a representatividade de cada palavra dentro do tópico em questão. A definição da LDA como *modelo probabilístico gerativo* assume que esses tópicos são preexistentes em relação à constituição dos documentos e são fixos no tempo. Essa pressuposição é necessária para que o processo de geração do documento representado pelo modelo seja possível (Blei, 2012), mas pode ser flexibilizada, permitindo, por exemplo, que os tópicos variem com o tempo.

Considerando  $K$  como um número fixo de tópicos,  $V$  o vocabulário do *corpus*, composto pelos termos que aparecem em quaisquer documentos,  $N$  o número total de palavras presentes num documento,  $D$  o número de documentos no *corpus*,  $\vec{\alpha}$  um vetor de dimensão  $K$ , e  $\eta$  um escalar, definimos  $\text{Dir}_K(\vec{\alpha})$  como uma distribuição de Dirichlet com  $K$  dimensões e parâmetro de concentração  $\vec{\alpha}$ ; e  $\text{Dir}_V(\eta)$  como uma distribuição de Dirichlet simétrica de  $V$  dimensões e parâmetro de concentração  $\eta$ . O processo de geração de

<sup>3</sup>Tecnicamente, essa distribuição multinomial se refere a apenas um ensaio, e portanto seria mais corretamente definida como uma distribuição categórica, conforme explicitado nas Equação 2.13 e Equação 2.14, acima. Porém, como não há nenhuma diferença entre uma distribuição categórica e uma distribuição multinomial com um ensaio, mantemos o termo de uso tradicional.



um documento pode então ser descrito da seguinte maneira (Blei & Lafferty, 2009):

1. Para cada tópico  $k \in \{1, \dots, K\}$ ,
  - a) Sortear uma distribuição sobre termos  $\vec{\beta}_k \sim \text{Dir}_V(\eta)$ .
2. Para cada documento  $\vec{w}_d \in \mathcal{D}$ ,
  - a) Sortear um vetor de proporções de tópicos  $\vec{\theta}_d \sim \text{Dir}_K(\vec{\alpha})$ .
  - b) Para cada palavra  $w_{d,n} \in \vec{w}_d$ ,
    - i. Sortear uma atribuição de tópico  $z_{d,n} \sim \text{Mult}(\vec{\theta}_d)$ ,  $z_{d,n} \in \{1, \dots, K\}$ .
    - ii. Sortear uma palavra  $w_{d,n} \sim \text{Mult}(\vec{\beta}_{z_{d,n}})$ ,  $w_{d,n} \in \{1, \dots, V\}$ .

Na versão original do modelo, Blei et al. (2003) consideram ainda o sorteio do número de palavras a partir de uma distribuição de Poisson com parâmetro  $\lambda$  fixo e estimado a partir dos dados. Como esse parâmetro não interfere na inferência dos outros parâmetros do modelo, ele foi deixado de lado nas definições posteriores.

A intuição formalizada no modelo é simples: cada documento abordará diferentes proporções de um número finito de tópicos; cada tópico é representado por um conjunto finito de palavras, e o documento é formado pela escolha de palavras a partir de diferentes tópicos. O modelo não leva em consideração a ordem das palavras, tratando cada documento como um *saco de palavras* (*bag of words*). Apesar de bastante ingênuo, o modelo reflete com fidedignidade a ideia de que um texto é composto a partir da abordagem de vários temas, sendo esses temas formulados por meio de diferentes palavras semanticamente próximas.

Como a LDA assume que um documento é composto por mais de um tópico, ela pode ser considerada um modelo de filiação mista (*mixed-membership model*), uma generalização dos modelos de mistura que só permitem a designação de um único tópico a cada documento. A capacidade de capturar a heterogeneidade de tópicos num documento amplia a generalização dos resultados alcançados pela LDA, superando o desempenho de modelos mais simples baseado na atribuição de um tópico único.

Para gerar a distribuição multinomial de proporções de tópicos por documento, definida na primeira etapa do modelo, uma distribuição de Dirichlet proposta *a priori* é utilizada. Cada documento no *corpus* é derivado do mesmo conjunto total de tópicos;

porém, cada documento terá proporções diferentes para cada um dos tópicos.

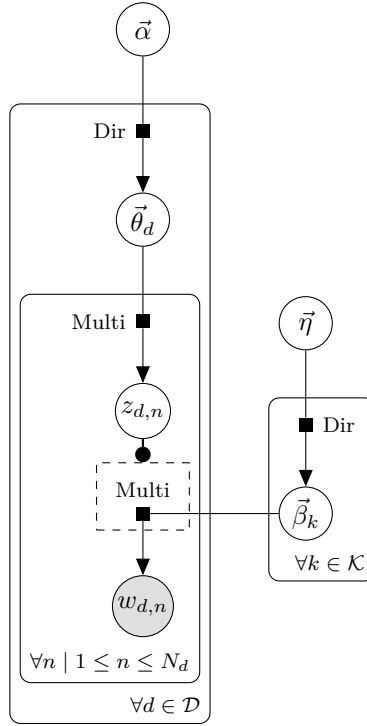
O modelo da LDA descobre a estrutura latente de tópicos revertendo o processo gerativo formalizado. Os documentos e as palavras que os constituem são o produto final ao qual se tem acesso, mas não há evidência explícita de quais foram os tópicos escolhidos para sua geração. A partir das informações observadas – ou seja, os padrões de coocorrência na distribuição de palavras no conjunto de documentos do *corpus* – o modelo infere a estrutura de tópicos e sua distribuição a partir do modelo gerativo (Griffiths & Steyvers, 2004).

O resultado da aplicação da LDA permite tanto a exploração da estrutura latente de tópicos presente num *corpus* quanto à categorização de documentos em diferentes classes, o que permite organizar os documentos, antes sem nenhuma estrutura, num banco de dados semanticamente significativo.

### 2.3.4 A LDA como modelo gráfico

Outra maneira de representar o modelo da LDA é por meio de um *modelo gráfico probabilístico*. As informações apresentadas pelo modelo gráfico não diferem da equação da probabilidade conjunta ou da descrição informal do modelo gerativo. Mesmo assim, a notação utilizada, denominada de notação de bandeja (*plate notation*), permite visualizar os parâmetros do modelo, suas replicações e dependências. A Figura 2.2 apresenta o modelo gráfico probabilístico para a LDA.

Cada vértice do grafo representa um dos parâmetros do modelo, já abordados acima. Os vértices em branco representam os parâmetros latentes que serão estimados. O vértice em cinza representa os dados observados – as palavras em cada documento. Cada replicação dos parâmetros aparece sobre uma bandeja na qual está especificado o tipo de distribuição da qual é retirado o parâmetro em questão. A distribuição da proporção de tópicos,  $\theta_d$ , é replicada para cada documento do *corpus* examinado, resultando num total de  $D$  parâmetros. Cada distribuição é aleatoriamente selecionada a partir de uma distribuição de Dirichlet definida *a priori* pelo parâmetro de concentração  $\vec{\alpha}$ . De maneira semelhante, cada tópico  $\beta_k$  é selecionado a partir de outra distribuição de Dirichlet, também definida *a priori*, parametrizada por  $\eta$ . Ambos os parâmetros  $\theta_d$  e  $\beta_k$  definem distribuições multinomiais que determinam a escolha



**Figura 2.2:** Modelo gráfico probabilístico da LDA (adaptado de Luttinen, 2012 e de Blei et al., 2003)

de tópicos para cada palavra num documento ( $z_{d,n}$ ) e a escolha de palavras específicas a partir do tópico escolhido ( $w_{d,n}$ ). Esses dois últimos parâmetros se encontram sobre a bandeja aninhada dentro da bandeja dos documentos.

O modelo gráfico ajuda a compreender porque este modelo de tópicos é chamado de *Alocação de Dirichlet Latente*. Tanto a distribuição da proporção de tópicos quanto os próprios tópicos são concebidos como o resultado de um processo que pode ser descrito por uma distribuição de Dirichlet. Em outras palavras, o resultado retirado da distribuição de Dirichlet serve para alocar as palavras dos documentos em tópicos latentes. A utilização do termo *latente* para designar a estrutura de tópicos e suas proporções nos documentos se deve aos desenvolvimentos da Análise Semântica Latente

(Landauer et al., 1998) e da pLSI (Hoffmann, 1999), descritas acima, na medida em que lidam com relações entre palavras que não estão explícitas nos documentos, mas que podem ser inferidas a partir dos padrões de coocorrência.

**Comparação com outros modelos similares** Além da LDA, outros modelos probabilísticos também representam aspectos do processo gerativo de documentos. Convém compará-los a partir de seus modelos gráficos para ressaltar as principais diferenças e indicar em quais aspectos a LDA é um avanço com relação a eles.

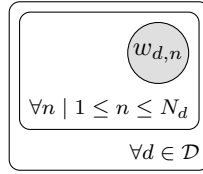
No modelo mais simples, o modelo de unigrama, cada palavra de todos os documentos são concebidas como provenientes de uma mesma distribuição multinomial. Inferência a partir da aplicação do modelo em um *corpus* resulta na distribuição multinomial dos termos mais frequentes – aqueles termos do vocabulário com maior probabilidade de aparecer em qualquer documento – o que por sua vez permite também computar a verossimilhança de um dado documento do *corpus*. A probabilidade de um documento no modelo pode ser definida da seguinte maneira (Blei et al., 2003):

$$p(\vec{w}) = \prod_{n=1}^N p(w_n) \quad (2.22)$$

Ou seja, o produto da probabilidade de cada palavra presente no documento. Como a probabilidade de cada palavra é definida por uma única distribuição multinomial, esse modelo acaba por superestimar a presença de termos frequentes e subestimar a frequência de ocorrência de termos mais raros, uma vez que ele considera sempre o *corpus* como um todo. Esse modelo também desconsidera as variações no uso do vocabulário em cada documento, utilizando uma distribuição comum a todos eles.

O modelo pode ser graficamente descrito com o seguinte grafo:

É possível incrementar o modelo de unigrama acrescentando uma variável de tópico por documento. O acréscimo dessa nova variável aleatória transforma o modelo de unigrama em uma *mistura* de unigramas. Isso significa que o *corpus* é compreendido como uma mistura de vários tópicos, mas cada documento é proveniente de um único tópico. O processo de geração de um documento sob esse modelo é dado da seguinte maneira: um tópico  $z$  é escolhido para cada documento e então, condicionado ao tópico escolhido,  $N$  palavras são geradas independentemente a partir da distribuição multinomial

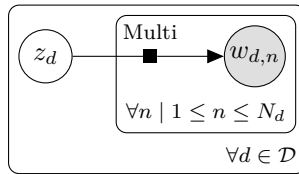


**Figura 2.3:** Modelo gráfico probabilístico do unigrama (adaptado de Blei et al., 2003)

$p(w | z)$ . A distribuição multinomial de termos correspondente a  $z$  pode ser interpretado como representação dos tópicos presentes no *corpus*. Esse modelo de mistura de unigramas nada mais é do que um classificador Bayesiano ingênuo (*Naive Bayes Classifier*) aplicado à análise textual. A representação formal da probabilidade de um documento sob esse modelo pode ser assim descrita:

$$p(\vec{w}) = \sum_z p(z) \prod_{n=1}^N p(w_n | z) \quad (2.23)$$

A representação gráfica do modelo, por sua vez, pode ser definida com o seguinte DAG:



**Figura 2.4:** Modelo gráfico probabilístico para a mistura de unigramas (Classificador Bayesiano Ingênuo) (adaptado de Blei et al., 2003)

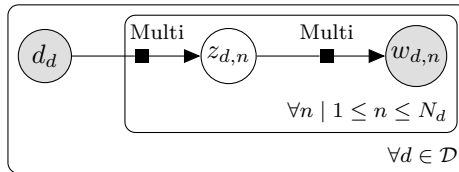
Em contraste com a LDA, a mistura de unigramas é limitada por atribuir apenas um tópico a cada documento. Isso torna pouco informativa a modelagem de grandes coleções de documentos com conteúdo heterogêneo.

Por fim, outro modelo gerativo utilizado para coleções de documentos é a *Indexação Semântica Latente Probabilística* (pLSI),

que amplia a proposta da mistura de unigramas relaxando o pressuposto de que cada documento contém apenas um único tópico. Na pLSI, cada documento é concebido como composto por diferentes proporções dos tópicos da coleção, ou seja, cada documento  $d$  é formado por uma mistura de distribuições sobre palavras  $z$ , e cada uma das  $N$  palavras do documento provém de atribuições de tópico próprias. A probabilidade conjunta de um documento  $d$  e de uma palavra  $w_n$  segundo esse modelo é dada pela seguinte equação:

$$p(d, w_n) = p(d) \sum_z p(w_n | z) p(z | d) \quad (2.24)$$

O problema da pLSI é que a variável  $d$  é apenas o índice do documento no *corpus*, o que significa que ela é uma variável multinomial com tantos valores possíveis quanto documentos existentes na coleção analisada. Em função disso, a pLSI não fornece um modelo gerativo completo para documentos, e as proporções que ela aprende a partir de um *corpus* são limitadas aos documentos presentes naquele *corpus*, impedindo uma verdadeira generalização para novos documentos. Outro resultado dessa limitação é o possível superajuste do modelo, por não haver nenhuma regularização no ajuste de  $p(d)$ . É por esse motivo também que não é possível modelar diretamente a probabilidade do documento  $d$  (ou  $\vec{w}$ ) a partir do modelo, pois ele é assumido como fixo. O DAG para a pLSI é definido abaixo:



**Figura 2.5:** Modelo gráfico probabilístico da pLSI (adaptado de Blei et al., 2003)

Para resolver essa limitação, a LDA trata a mistura de tópicos como uma variável aleatória latente com  $k$  parâmetros, abrindo mão de modelar cada documento específico para definir uma tendência central de geração de proporções. Além de diminuir o número de variáveis a serem ajustadas no modelo, essa estratégia também previne o superajuste e facilita a predição de novos documentos.

### 2.3.5 A LDA como modelo probabilístico

A LDA é um exemplo de *modelagem probabilística* (Blei, 2012). A modelagem probabilística gerativa teoriza que os dados a serem analisados emergem de um processo que, mesmo tendo flutuações aleatórias, pode ser explicado por variáveis latentes. Por meio do modelo proposto, é possível estimar essas variáveis latentes a partir dos dados observados. Nesse sentido, a modelagem probabilística serve-se das definições utilizadas acima para caracterizar a inferência bayesiana em modelos gráficos.

A definição formal do processo gerativo é feita por meio de uma distribuição de probabilidade conjunta para as variáveis latentes e os dados observados. A distribuição de probabilidade conjunta permite calcular a distribuição condicional das variáveis latentes em função das informações observadas. A distribuição das variáveis latentes condicionada aos dados observados, por sua vez, configura uma *distribuição posterior* no quadro da inferência bayesiana.

No caso da LDA, os dados observados são as palavras distribuídas nos documentos; as variáveis latentes são as estruturas de tópicos que o modelo assume existir e determinar a constituição dos documentos por meio do processo gerativo.

Para representar de maneira formal o processo gerativo modelado pela LDA, utilizaremos a notação empregada acima para abordar a definição da distribuição conjunta do modelo para inferência bayesiana. Os tópicos que serão inferidos a partir dos dados são representado por  $\beta_{1:K}$ , sendo que cada  $\beta_k$  representa uma distribuição multinomial que atribui diferentes probabilidades aos termos presentes no vocabulário. A proporção de tópicos para um documento específico é representada por  $\theta_{d,k}$ : a proporção do tópico  $k$  dentro do documento  $d$ . A atribuição de tópicos às palavras dos documentos é representada por  $z_{d,n}$ : a  $n$ -ésima palavra do documento  $d$  pertence ao tópico  $z$ . Por fim, temos os dados efetivamente observados representados por  $w_{d,n}$ : a palavra  $w$ , pertencente ao vocabulário estático composto a partir de todos os documentos, é o  $n$ -ésimo elemento do documento  $\vec{w}_d$ . Com base nessa notação, a distribuição conjunta da LDA pode ser assim definida (Blei, 2012), se suprimirmos os hiperparâmetros:

$$p(\beta_{1:K}, \theta_{1:D}, \vec{z}_{1:D}, \vec{w}_{1:D}) = \prod_{i=1}^K p(\beta_i) \prod_{d=1}^D p(\theta_d) \left[ \prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right] \quad (2.25)$$

A probabilidade conjunta de todos os parâmetros de interesse (os tópicos  $\beta_{1:K}$ , as proporções de tópicos por documento  $\theta_{1:D}$ , a atribuição de tópicos para cada palavra do documento  $\tilde{z}_{1:D}$  e cada palavra presente nos documentos  $\tilde{w}_{1:D}$ ) é igual (lendo a equação de trás para frente) ao produto da probabilidade da atribuição de tópico a cada palavras em cada documento, dada a proporção de tópicos em cada documento, multiplicada pela probabilidade de cada palavra em cada documento, dada a distribuição de todos os tópicos em conjunto com a atribuição de tópicos de cada palavra para cada documento. Esse produto, por sua vez, é multiplicado pela probabilidade da distribuição de tópicos em cada documento (aqui implicitamente condicionada a  $\tilde{\alpha}$ ), que por fim é multiplicado pelo produto da probabilidade de cada tópico (também implicitamente condicionado a  $\eta$ ).

As dependências apresentadas na equação acima definem a especificidade da LDA. O modelo teórico deixa evidente que a atribuição de tópicos às palavras em um dado documento ( $z_{d,n}$ ) depende da proporção de tópicos para aquele documento ( $\theta_d$ ). A probabilidade de cada palavra em um dado documento ( $w_{d,n}$ ) depende da atribuição dos tópicos às palavras ( $z_{d,n}$ ) e da probabilidade de todos os tópicos ( $\beta_{1:K}$ ). Outra maneira de entender essa última dependência é pensar que, para determinar a probabilidade de uma palavra dentro de um documento, o modelo determina a qual tópico aquela palavra está atribuída ( $z_{d,n}$ ) e retorna a probabilidade daquela palavra no tópico de atribuição ( $p(w|\beta)$ ).

Podemos compreender a distribuição conjunta formalmente definida pela Equação 2.25 acima relacionando-a com a descrição informal do processo gerativo hipotetizado pelo modelo. O primeiro termo do lado direito da equação ( $\prod_{i=1}^K p(\beta_i)$ ) refere-se ao conjunto de tópicos que o modelo supõe como preexistente à formação dos documentos. Como a equação trata da probabilidade conjunta de todos os parâmetros, e assume cada tópico como independente um do outro, a probabilidade conjunta dos tópicos é simplesmente o produto da sequência de todos os tópicos. O modelo concebe a elaboração de um documento primeiramente pela escolha de uma distribuição de diferentes tópicos. Esse aspecto, apresentado na etapa (2a) na descrição gerativa do modelo, é representado na equação pela primeira parte do segundo termo ( $\prod_{d=1}^D p(\theta_d)$ ). Com uma proporção de tópicos já determinada, o modelo assume que o documento é criado a partir da repetida escolha de tópicos dentro da distribuição da proporção de tópicos. Esse aspecto, descrito na



etapa (2b.i) do modelo gerativo, é representado na equação pelo termo  $\prod_{n=1}^N p(z_{d,n}|\theta_d)$ . Por fim, tendo escolhido um tópico, escolhemos uma palavra dentro do tópico. Esse processo final, que gera os dados aos quais temos acesso, é descrito pela etapa (2b.ii), acima, e pelo termo  $\prod_{n=1}^N p(w_{d,n}|\beta_{1:K}, z_{d,n})$ .

De maneira geral, em consonância com as práticas da estatística bayesiana, o parâmetro de concentração das distribuições de Dirichlet *a priori* do modelo é fracamente informativo, o que significa que, *a priori*, cada tópico tem igual probabilidade de aparecer em um documento e cada palavra do vocabulário tem igual probabilidade *a priori* de ser representativa de qualquer tópico. As efetivas probabilidades são inferidas a partir dos dados. Porém, como notam Wallach, Minmo, e McCallum (2009), o uso de parâmetros de concentração simétricos ou assimétricos, bem como o valor desses parâmetros, tem forte influência sobre o ajuste do modelo.

### 2.3.6 Estimativas dos parâmetros da LDA

A estimativa dos diversos parâmetros latentes da LDA é calculada por meio de inferência bayesiana, invertendo o processo gerativo hipotetizado pelo modelo. Partindo da definição da distribuição de probabilidade conjunta, o modelo da LDA calcula aproximadamente a distribuição posterior dos parâmetros latentes condicionados aos dados observados. Com base no teorema de Bayes, a distribuição posterior pode ser representada da seguinte maneira, utilizando a notação da equação da distribuição conjunta dos parâmetros (Blei, 2012):

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}|\vec{w}_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, \vec{w}_{1:D})}{p(\vec{w}_{1:D})} \quad (2.26)$$

A *distribuição posterior*, ou seja, a probabilidade dos parâmetros latentes condicionada aos dados observados (o termo à esquerda da equação), é igual à *distribuição conjunta* de todos os parâmetros do modelo – tradicionalmente representada no teorema de Bayes como o produto da *função de verossimilhança* com a *distribuição a priori* – dividida pela *evidência*, ou seja, a distribuição marginal dos dados observados.

A grande dificuldade em calcular diretamente a distribuição posterior é a impossibilidade de resolver analiticamente o denominador da equação. O cálculo da evidência depende da soma de todas

as possíveis variações da distribuição de tópicos, feito impossível mesmo com um pequeno número de documentos e tópicos. Para resolver este impasse, a LDA faz uso de métodos aproximativos para calcular a distribuição posterior.

O primeiro desses métodos, amplamente utilizado em outras áreas que fazem uso de inferência bayesiana, é a aproximação por amostragem da distribuição posterior utilizando Monte Carlo por Cadeia de Markov (*Markov Chain Monte Carlo*, MCMC), abordada acima na Subsubseção 2.1.4.2. Em particular, a *amostragem de Gibbs* se revela particularmente útil para calcular os parâmetros da LDA, graças à conjugalidade entre a distribuição de Dirichlet e a distribuição multinomial.

O segundo método, fruto de desenvolvimentos mais recentes das pesquisas em estatística bayesiana, é a aproximação denominada de método variacional (Blei, 2012). Em vez de simular amostras aleatórias da distribuição posterior, os métodos variacionais propõem uma família de distribuições aproximativas com parâmetros computáveis e busca, por otimização, encontrar a distribuição dentre as propostas que mais se aproxima da distribuição posterior, diminuindo a divergência de Kullback-Leibler, conforme descrevemos em maior detalhe na Subsubseção 2.1.4.3.

A opção por um ou outro método de aproximação da distribuição posterior é contingente ao problema em questão e aos recursos computacionais disponíveis. De maneira geral, a aplicação dos diferentes algoritmos produzem resultados bastante semelhantes. Para o problema específico proposto pela LDA, já existem implementações dos diferentes tipos de aproximação, o que deixa a escolha dos métodos computacionais à preferência do pesquisador (Blei, 2013).

### 2.3.7 Desenvolvimento de outros modelos de tópicos

Uma das maiores vantagens da LDA, resultante de sua definição como modelo probabilístico, é seu caráter modular. Ela pode ser facilmente ampliada ou transformada, incluída dentro de outros modelos mais amplos ou mais complexos. Por esse motivo, o trabalho original da LDA (Blei et al., 2003) foi seminal e ampliou significativamente as pesquisas sobre o tema. O trabalho prático de implementação da LDA realizado por Griffiths e Steyvers (2004) demonstrou, pouco depois da proposição do modelo original, a potência da LDA para extrair informações semânticas de coleções de documentos –

nesse caso, encontrar os temas de diferentes artigos a partir da coleção de resumos da *Proceedings of the National Academy of Science*.

Os modelos de tópicos desenvolvidos sob influência da LDA adaptam sua formulação simples para permitir relaxar os pressupostos mais básicos, como a independência dos tópicos entre si e o caráter não informativo da ordem das palavras no documento; ou então adicionar dados externos ao documento, como metadados. O desenvolvimento de modelos de tópicos com pressupostos mais flexíveis é uma área de pesquisa bastante ativa (Blei, 2012). O principal objetivo em relaxar os pressupostos da LDA é obter modelos que retratem de maneira mais verossímil a composição dos documentos de uma coleção – levando em consideração a ordem das palavras, por exemplo – e descobrir estruturas mais complexas, como tópicos correlacionados e estruturas hierárquicas de temas.

Um dos principais pressupostos da LDA é a noção de *saco de palavras* (*bag of words*), ao qual já fizemos referência, comum a outros modelos de linguagem natural. O modelo de saco de palavras não leva em consideração a ordem das palavras ou a estrutura gramatical de um documento. Para esse pressuposto, um documento não passa de um punhado de palavras agrupado num mesmo conjunto. Apesar de esse pressuposto ferir de maneira evidente qualquer entendimento intuitivo sobre o funcionamento da linguagem, ainda assim consegue revelar a estrutura latente de tópicos pela co-ocorrência de termos num mesmo documento. De fato, apesar de ser uma pressuposição extremamente simplista para dados linguísticos, seu desempenho em diferentes tarefas de análise de dados textuais é geralmente próximo do desempenho de modelos muito mais concretos, que levam em consideração a classe sintática e morfológica das palavras, em função dos modelos mais complexos serem fortemente afetados pelo problema da alta dimensionalidade dos dados.

Alguns modelos mais recentes expandem a LDA conjugando-a com um Modelo Oculto de Markov (*Hidden Markov Model*, HMM) (Andrews & Vigliocco, 2010). O modelo resultante condiciona a escolha de tópicos e palavras na formação de um documento às palavras imediatamente anteriores, assumindo que a probabilidade de uma palavra ser escolhida é dependente das palavras que vieram antes dela. Apesar do modelo não fazer uso de informações sintáticas explícitas, levar em consideração as palavras antecedentes permite modelar de maneira mais fidedigna o processo gerativo da linguagem.

Outra implementação que conjuga a LDA com um Modelo

Oculto de Markov, desenvolvido por Griffiths, Steyvers, Blei, e Tenenbaum (2004), consegue distinguir entre palavras com função sintática e palavras com função semântica numa sentença. Utilizando uma variável de estado oculto que indica se uma palavra cumpre função sintática ou semântica, o modelo consegue inferir, sem nenhum tipo de supervisão, palavras com conteúdo semântico mais significativo, organizando-as em tópicos.

O modelo de tópico estrutural (strTM) conjuga a habilidade de distinguir entre palavras funcionais e semânticas com a dependência da transição de um tópico para outro com base no tópico precedente, ou seja, assume que os tópicos se seguem como uma cadeia de Markov de primeira ordem (H. Wang, Zhang, & Zhai, 2011). Esse modelo também amplia a noção de tópico aplicando-o às orações do documento, em vez de cada palavra tomada isoladamente.

Outro pressuposto da LDA é que a ordem dos documentos não tem nenhuma importância – consequência derivada da independência assumida entre os elementos componentes. Assim como cada documento é modelado como um saco de palavras, o *corpus* é entendido como um saco de documentos. Esse pressuposto pode ser bastante aceitável quando o modelo é aplicado a um conjunto de documentos nos quais não esperamos nenhuma ou pouca influência do tempo; ou então quando não é necessário assumir que um documento influencia na formação de outro. O maior problema desse pressuposto é quando há interesse em se analisar as transformações no conteúdo e proporções de tópicos numa grande série temporal, por exemplo. Para endereçar essa limitação, foram propostos modelos de tópicos dinâmicos (Blei & Lafferty, 2006; X. Wang & McCallum, 2006). A grande diferença desses modelos com relação à LDA é a concepção de tópico como uma sequência de distribuições de palavras, e não mais uma única distribuição multinomial do vocabulário. Ao levar em consideração a cronologia dos documentos, os modelos de tópicos dinâmicos permitem descrever uma estrutura mais complexa de tópicos e acompanhar as mudanças ocorridas nos tópicos ao longo do tempo.

O modelo da LDA também torna necessário definir *a priori* o número de tópicos presentes numa coleção de documentos. Ainda que uma análise informada possa prever com alguma margem de erro a quantidade de tópicos de um *corpus*, dificilmente podemos ter certeza desse número e, na maioria das vezes, seria preferível que a análise informasse o número mais provável de tópicos. Essa limitação da LDA é remediada pelo modelo de tópico bayesiano não-

paramétrico (Blei, 2012; Blei, Griffiths, & Jordan, 2010). Apesar do nome, esse modelo trabalha com distribuições parametrizadas, mas a distribuição posterior é calculada como sendo composta por um número de parâmetros que pode aumentar em função dos dados e ser potencialmente infinita. As dificuldades computacionais implicadas pelo modelo são compensadas pela riqueza da estrutura latente descoberta: além de inferir o número mais provável de tópicos a partir dos dados observados, o modelo não-paramétrico pode ser estendido para inferir uma árvore hierárquica de tópicos, partindo de tópicos mais gerais e abrangentes para outros mais específicos.

Outro conjunto de modelos que expandem os pressupostos da LDA são aqueles que assumem a existência de correlações entre os tópicos, como o modelo de tópicos correlacionados e a máquina de alocação de Pachinko (Blei & Lafferty, 2007; Li & McCallum, 2008). A inferência da correlação entre tópicos permite elaborar visualizações dos tópicos de uma coleção sob a forma de grafos, com a distância entre os vértices indicando a proximidade entre um tópico e outro. Alterações na máquina de alocação de Pachinko ou a utilização do processo do restaurante chinês permitem ampliar a LDA para considerar não só correlações entre tópicos, mas também uma estrutura hierárquica entre eles, partindo de tópicos mais genéricos para tópicos específicos aninhados (Blei et al., 2010).

Há modelos de tópicos que não necessariamente expandem os pressupostos da LDA, mas modificam o modelo probabilístico de forma a permitir que dados exteriores aos documentos – metadados – possam ser levados em consideração na inferência da estrutura latente. Um desses modelos, por exemplo, permite associar tópicos com autores específicos. Essa adição dos dados de autoria permite verificar a similaridade entre autores com base na proporção de tópicos utilizados, ao mesmo tempo em que mantém a habilidade da LDA em verificar os tópicos dos documentos. Roberts et al. (2014), na proposta dos *Structural Topic Models* (STM), ampliam a flexibilidade de adição de metadados, permitindo incluir covariáveis categoriais e numéricas que alteram a proporção de tópicos prevista nos documentos, conforme apresentaremos em maiores detalhes na implementação prática da estimativa de nosso modelo, na Seção 3.3.

## 2.4 Utilização dos modelos de tópicos na Psicologia

Mesmo sendo um tema em pleno desenvolvimento e pesquisa em sua área de origem, os modelos de tópicos já são utilizados em pesquisas de diferentes campos de conhecimento que lidam com dados textuais. Muitos dos modelos que estendem os pressupostos da LDA foram desenvolvidos por pesquisadores que precisavam endereçar problemas específicos de sua área de pesquisa (Blei, 2012). Os modelos de tópicos podem servir como uma ferramenta para auxiliar pesquisadores das ciências humanas que lidam com dados textuais de diversas origens (Blei, 2013). De fato, uma das áreas onde a LDA encontrou grande repercussão prática foi as Humanidades Digitais (*Digital Humanities*), especialmente como aplicação da estratégia de *leitura distante* de documentos históricos e obras literárias ou para o trabalho quantitativo com respostas abertas em pesquisas nas Ciências Sociais (Roberts, Stewart, Tingley, & Airoldi, 2013).

Blei (2013) propõe, porém, que os pesquisadores das áreas das humanidades com interesse no uso de tecnologias digitais como ferramenta de pesquisa não se limitem a aplicar os modelos desenvolvidos por engenheiros informáticos, mas que façam uso da modelagem probabilística para representar estruturas latentes que reflitam de maneira significativa os conhecimentos teóricos sobre sua área de investigação. O uso de modelos probabilísticos permitiria às diferentes áreas das humanidades, historicamente marcada pela abordagem qualitativa de textos, elaborar modelos formais que poderiam ser testados, ampliados e refutados de acordo com sua adequação aos dados disponíveis.

Blei (2012) também adverte sobre os riscos da utilização dos modelos de tópicos em análise exploratória em diferentes campos científicos. Em sua origem, os modelos de tópicos são concebidos como uma ferramenta para facilitar a recuperação de informação em grandes coleções de documentos, e portanto não se preocupam com a validade teórica dos resultados encontrados. Por sua natureza exploratória, os trabalhos de aplicação dos modelos de tópicos na área das Ciências da Computação não se preocupam em resolver diretamente o problema da *multimodalidade* que pode afetar qualquer modelo de mistura com variáveis latentes, como a LDA e modelos derivados (Roberts, Stewart, & Tingley, 2016). A utilização de modelos de tópicos na investigação científica, porém, traz consigo a exigência de estratégias de validação dos resultados. Essas estraté-

gias precisam ser pensadas a partir de cada aplicação particular e cada campo do conhecimento.

No campo da Psicologia, em particular, já são desenvolvidas diversas linhas de pesquisa que implementam modelos de tópico para modelar fenômenos de interesse e investigar dados textuais. A área que mais prontamente utilizou os modelos de tópicos foi a Psicologia Cognitiva, influenciada pela ampliação das pesquisas de modelos bayesianos para os processos cognitivos (Griffiths et al., 2008; Griffiths, Steyvers, & Tenenbaum, 2007). Uma vez que modelos probabilísticos já eram utilizados para representar diversos processos cognitivos, a LDA e os modelos de tópicos foram rapidamente adotados, principalmente por pesquisadores interessados em modelar representações semânticas e processos de inferência de significado a partir de dados de contexto. Os resultados obtidos pela comparação da LDA com outros modelos de representação semântica apontam para um melhor desempenho do modelo de tópicos em prever uma série de atividades cognitivas relacionadas à memória semântica em sujeitos humanos: associação de palavras (Socher et al., 2009), desambiguação de palavras homófonas, extração de tema de um texto (Griffiths et al., 2007). Na Psicologia Cognitiva, portanto, os modelos de tópicos são utilizados principalmente como modelos teóricos para explicar fenômenos relativos ao processamento da linguagem em sua dimensão semântica (Jones, Willits, & Dennis, 2015).

Há outras áreas de pesquisa em Psicologia que utilizam os modelos de tópicos como ferramenta para análise exploratória de dados. Em particular, há grupos de pesquisa que exploram a utilidade dos modelos de tópicos para a investigação de dados textuais provenientes da prática clínica do psicólogo, num sentido próximo do que estamos propondo nesta tese. A quantidade de trabalhos publicados, entretanto, ainda é pequena, se considerarmos que o artigo sobre a LDA é de 2002. Além da utilidade exploratória, o fato de os modelos de tópico permitirem quantificar diversos aspectos ligados à dimensão semântica torna viável sua utilização para a predição de variáveis externas ao conteúdo do texto.

Isso abre caminho para dois desenvolvimentos importantes: primeiro, a verificação da validade de critério do modelo de tópico com relação a uma variável externa aos textos investigados; segundo, a ampliação do uso de dados textuais em modelos quantitativos, como regressões, fazendo uso dos coeficientes resultantes da aplicação do modelo de tópico como covariáveis na predição de uma variável resultante.

O primeiro trabalho de aplicação de modelos de tópicos para a análise de dados de psicoterapias é a tese de doutorado de Michelle Doeden (2010), defendida em 2010 e orientada por David Atkins. Doeden fez uso da LDA para explorar um *corpus* formado pela transcrição integral de uma seleção de quatro sessões de psicoterapia de 91 casais que foram submetidos ou à Psicoterapia Comportamental de Casal Tradicional ou à Psicoterapia Comportamental de Casal Integrativa. O uso dos modelos de tópicos, porém, permaneceu restrito ao papel de uma ferramenta para auxiliar na investigação de um grande *corpus*, sem discussões sobre a pertinência do modelo para os dados analisados ou qualquer relação entre as teorias de fundamento das psicoterapias e os pressupostos dos modelos. Mesmo assim, Doeden avaliou os resultados de um modelo com 100 tópicos em termos de seu desempenho em prever, por meio de modelos de regressão logística utilizando os tópicos como preditores, se o enunciador era cliente ou psicoterapeuta, ou se era homem ou mulher; qual era o tipo de psicoterapia e qual o nível de satisfação com o relacionamento. Os modelos de regressão indicaram que os dados dos tópicos permitiram distinguir os turnos de falas pertencentes ao terapeuta e aos clientes, mas não entre homem e mulher. A distinção entre tipo de psicoterapia e tipo de satisfação com o relacionamento também foi bem previsto pelo modelo.

Os resultados da pesquisa de Doeden são apresentados e ampliados em um artigo de seu orientador (Atkins et al., 2012). Neste artigo também são apresentados modelos de regressão logística para prever as repostas aos itens da escala *Couple-Therapy in-Session Behavior-Rating System*, que é dividida em quatro subescalas: habilidades de resolução de conflito construtivas; discussão descritiva; emoção positiva; e culpabilização. Avaliando os modelos nos dados de treinamento e por meio de validação cruzada, os autores verificaram que as informações sobre os tópicos dão aos modelos preditivos uma acurácia consideravelmente maior do que o esperado pelo acaso, variando entre 55% e 85%. Com isso, concluem que os modelos de tópicos podem ser uma ferramenta útil para automatizar a classificação de dados de transcrições de psicoterapias, permitindo realizar estudos com maior volume de dados. Da mesma maneira, esperamos que nossa proposta de um modelo de tópicos para as associações livres permita analisar grandes conjuntos de dados provenientes das sessões clínicas de Psicanálise, facilitando a comunicação de aspectos relevantes e diminuindo o viés do pesquisador na seleção e interpretação dos resultados.



Em outro trabalho do grupo de Atkins (Atkins, Steyvers, Imel, & Smyth, 2014), os modelos de tópicos foram utilizados para extrair informações semânticas de um *corpus* da transcrição gravação de 148 Entrevistas Motivacionais (*Motivational Interviewing, MI*) conduzidas com usuários de álcool e outras drogas. O principal objetivo do estudo foi comparar os tópicos encontrados pela análise automática produzida pelo modelo de tópicos supervisionado com os códigos aplicados às mesmas entrevistas por avaliadores humanos. A codificação das sessões foi feita por meio do instrumento *Motivational Interviewing Skilled Code* (MISC), que serve para avaliar a fidedignidade com que cada terapeuta implementa a metodologia do MI, e também o quanto os pacientes relatam as alterações de seus comportamentos. Os resultados para a predição dos códigos foi consideravelmente alta, com uma área sob a curva ROC de 0,72. Porém, comparando a concordância da codificação feita por pessoas e a codificação feita pelo modelo, os resultados variavam consideravelmente, com o modelo obtendo desempenho próximo ao humano para sessões inteiras, mas com os humanos tendo maior concordância na classificação de turnos de falas.

Em trabalho semelhante, Gaut, Steyvers, Imel, Atkins, e Smyth (2015) utilizaram o modelo de tópicos L-LDA (*Labeled Latent Dirichlet Allocation*) para descobrir associações entre tópicos de palavras e códigos de metadados associados às transcrições de sessões de psicoterapia de diversas modalidades teóricas. As inferências feitas a partir da L-LDA foram comparadas com modelos mais simples, como regressão linear utilizando *lasso*, uma técnica de regularização dos coeficientes de regressão para reduzir a variância das estimativas. Os resultados indicam que o modelo de tópicos tem performance classificatória superior ao modelo de regressão, com área sob a curva ROC de 0,79 para a predição de rótulos relativos às sessões completas. No caso dos rótulos relativos aos turnos de fala, o desempenho dos dois modelos é semelhante – em concordância com o resultado do estudo de Atkins.

Por fim, Imel, Steyvers, e Atkins (2015a) fizeram uso de modelos de tópicos para explorar um grande *corpus* composto pela transcrição de 1.553 psicoterapias de diversas orientações e sessões médicas com psiquiatras. Diferente dos trabalhos anteriores, que avaliaram os resultados dos modelos de tópicos em função de algum padrão de referência, o artigo de Imel é principalmente de caráter exploratório, buscando identificar os principais agrupamentos de temas nas sessões de psicoterapia, além de diferenciar as modali-

dades de psicoterapias por meio dos tipos de tópicos utilizados nas intervenções dos terapeutas. A diferenciação entre técnicas psicoterápicas foi avaliada pela técnica de *random forests*, classificando as sessões de psicoterapia com base nos tópicos descobertos, e seu resultado foi comparado com as classificações originais do *corpus*. O erro em validação cruzada foi de 13%, indicando um bom poder preditivo das classificações geradas automaticamente pelos modelos e tópicos e sugerindo uma diferença nos temas e estilos de intervenções utilizados por diferentes psicoterapias.

### 2.4.1 O uso de modelos de tópicos para análise de associações livres

Esta tese se encontra, de certa maneira, na esteira do trabalho de Doeden, Atkins, Gaut e Imel, uma vez que também tem o objetivo de aplicar modelos de tópicos para explorar dados provenientes da transcrição de sessões clínicas. Diferentemente dos referidos trabalhos, porém, não tomamos os modelos de tópicos somente como uma ferramenta de valor meramente heurístico, dissociado das teorias utilizadas para fundamentar as psicoterapias. Nossa proposta busca utilizar toda a potencialidade dos modelos de tópicos como modelos probabilísticos, com o objetivo de formalizar um modelo digital a partir de modelos analógicos psicanalíticos, em consonância com a sugestão de Blei para os pesquisadores da área das Ciências Sociais e Humanas.

A intenção do uso do modelo de tópico como formalização de um objeto teórico também aproxima este trabalho das várias pesquisas empreendidas por Griffiths et al. (2008) para compreender o funcionamento da memória semântica. Desviamos parcialmente de seu objeto de estudo, contudo, pois não estamos interessados num modelo geral para a memória semântica, mas num modelo formal específico para o processo de livre associação a partir de seus determinantes inconscientes.

Tendo um panorama geral da inferência bayesiana e da modelagem probabilística, e tendo descrito em detalhes o principal modelo de tópicos, a Alocação de Dirichlet Latente, podemos passar à retomada dos modelos analógicos sobre a associação livre para finalmente definirmos nosso objeto modelo – como representaremos as associações e os processos psíquicos que lhes dão origem – e sua descrição por meio de uma teoria específica, baseada na teoria geral da probabilidade. Também abordaremos em seguida os procedimen-

tos que serão utilizados para ajustar e avaliar o modelo proposto a partir de um estudo de caso.



## Capítulo 3

# Modelos e procedimentos

Nosso próximo passo, já tendo discutido os modelos analógicos propostos por Freud e a teoria geral dos modelos de tópicos, é elaborar um objeto modelo das associações livres. Um objeto modelo, por definição, é uma simplificação do fenômeno que queremos compreender, e isso implica em descrevermos o processo de livre associação de maneira idealizada e passível de ser compreendida a partir dos modelos de tópicos. A inserção desse objeto modelo na teoria dos modelos de tópicos permitirá derivar nossa *teoria específica* (Bunge, 1974) sobre as associações livres: a descrição digital, matemática, dos processos que as engendram e como essa descrição formal pode ser contrastada com dados obtidos na clínica psicanalítica.

Em função da singularidade de cada sujeito, o modelo aqui proposto não pode ser mais do que uma *classe abstrata* de modelos possíveis. Sua utilidade só poderá ser auferida na medida em que o modelo abstrato é instanciado a partir de uma análise específica. Por isso, avaliaremos a adequação do modelo proposto a partir de um *estudo de caso*. As inovações propostas nesta tese, no sentido do desenvolvimento de um modelo digital para as associações livres, portanto, são acompanhadas por um alinhamento à tradição psicanalítica de construir seu objeto empírico a partir de estudos de caso. Nossa proposta passa por assumir que um caso clínico pode ser reduzido, à título de análise, a um *corpus* textual cuja organização latente procuraremos inferir utilizando nosso modelo teórico. Portanto, conduziremos uma análise o mais próximo possível do *setting* tradicional – nosso laço com a prática psicanalítica proposta por Freud – mas analisaremos os dados a partir de nossa proposta original, utilizando modelos de tópicos.

### 3.1 Objeto modelo das associações livres

Como psicanalistas, não podemos nos furtar do problema que é aproximar uma experiência singular, como a análise, de um modelo geral. Temos de nosso lado a pretensão, afirmada e repisada por Freud, de que a Psicanálise é uma ciência no sentido estrito, e, por isso, deve ter por objetivo construir modelos gerais e generalizáveis da experiência psíquica. Elegemos, dentre as várias teorias locais existentes na Psicanálise, as livre associações como foco de nosso estudo. Afinal, toda análise pessoal deve, por definição, fazer uso das associações livres como método de investigação; ademais, os casos sobre os quais Freud construiu a metapsicologia foram quase todos conduzidos sob o uso da regra fundamental da Psicanálise. Propondo um objeto modelo das associações livres, garantimos que nosso modelo geral de fato tem um referente concreto localizável na realidade e não é uma abstração pura.

Utilizando a notação desenvolvida por Bunge (1974), nosso referente  $R$  é delimitado pelo tipo específico de comunicação que acontece em sessões clínicas de Psicanálise. O analisando, a pedido do psicanalista, fala livremente o que lhe ocorre; o psicanalista, por sua vez, escuta de maneira equiflutuante e sem direcionar intencionalmente os fluxos associativos do analisando. Esse referente precisa ser abstraído numa classe de equivalência  $S$ , na qual representamos os aspectos essenciais do processo de livre associação, generalizáveis para todos os casos, e deixamos de lado aspectos periféricos que, apesar de poderem ser relevantes, não são de consequência grave para nossa formalização e podem, a princípio, ser ignorados. Essa classe de equivalência é definida pelos seguintes predicados, que serão desenvolvidos logo abaixo:

- $P_1$ – Uma associação livre tem como elemento mínimo um *termo*, instanciado como uma *palavra*;
- $P_2$ – Um conjunto de associações livres é delimitado por um *turno de fala*;
- $P_3$ – As intervenções do analista são extensões das associações livres do analisando por fazerem referência a seus determinantes.
- $P_4$ – A ordem em que os termos surgem numa associação livre depende apenas da sintaxe da língua e não é relevante para seu conteúdo;

- $P_5$ – Elementos não-verbais, tais como gesticulações e prosódia, também não são relevantes para o conteúdo das associações livres;
- $P_6$ – O fluxo de associações é determinado estocasticamente por *complexos associativos latentes*;
- $P_7$ – Os complexos associativos são formados por *conjuntos de termos* semanticamente associados;
- $P_8$ – Os complexos associativos são investidos de *libido* de acordo com sua importância na vida psíquica do sujeito;
- $P_9$ – A organização dos termos de um complexo é relativamente estável, mas a libido pode se deslocar com relativa facilidade entre complexos.

Alguns desses predicados –  $P_1$ ,  $P_2$  e  $P_3$  – são razoavelmente autoevidentes a partir do fato de que numa análise, nas palavras de Freud, duas pessoas conversam – ainda que sob regras bastante específicas para permitir separar as associações livres do fenômeno do diálogo. Chamaremos esses predicados de *predicados de caracterização*. Outros predicados –  $P_4$  e  $P_5$  – implicam em simplificações que tornam o objeto modelo mais prontamente generalizável e interpretável a partir dos modelos de tópicos. Esses predicados serão denominados de *predicados de delimitação*. Por fim, alguns predicados –  $P_6$ ,  $P_7$ ,  $P_8$  e  $P_9$  – estabelecem a característica dos processos latentes que causam a emergência das associações em si. Denominaremos esse conjunto de predicados de *predicados de determinação*.

### 3.1.1 Predicados de caracterização

Os predicados de caracterização,  $P_1$ ,  $P_2$  e  $P_3$ , estabelecem as características positivas das associações livres que serão posteriormente descritas pelo modelo teórico. Um primeiro ponto que precisa ser esclarecido é o que estamos entendendo por um *termo*, elemento constitutivo das associações. Um *termo* é um conceito representado pela forma canônica de uma palavra. Mais precisamente, um termo é o conjunto de todas as inflexões de um mesmo lexema, compartilhando um mesmo significado, portanto, e representado formalmente por um lema.

Obviamente, em função das exigências da língua, um termo só aparece na fala instanciado como uma palavra. Na instanciação,

o termo pode ser modificado para que possa dar conta da regência e concordância exigidas pela língua, além de permitir pequenas variações semânticas, como na temporalidade da conjugação de um verbo ou no gênero ou número na inflexão de um substantivo. O ponto principal que nos interessa para definir o elemento constitutivo de uma associação, entretanto, se foca no termo e não em sua instância concreta, na medida em que, para nosso objeto modelo, as inflexões nas instâncias concretas de um termo são acidentes exigidos pela língua mas de forma alguma necessárias do ponto de vista do processo gerador das associações livres.

A caracterização da unidade mínima das associações livres é razoavelmente autoevidente e só precisa ser explicitada no sentido de diferenciar os termos de suas instanciações em palavras. O conjunto máximo, porém, é necessariamente arbitrário. Afinal, qual é o ponto de corte ideal para podermos definir que uma determinada cadeia associativa finalizou e outra se iniciou? É da natureza das cadeias conterem os elementos de transição entre dois extremos, a tal ponto de ser quase impossível distinguir onde exatamente uma termina e outra inicia. Frente a essa dificuldade, resta-nos operar um corte arbitrário: um conjunto de associações relacionadas, mas cuja cardinalidade desconhecemos, é delimitado por um turno de fala.

Utilizamos o termo *turno de fala* no sentido mais fenomênico possível. Como numa análise há apenas dois interlocutores, o analisando e o analista, um turno de fala termina quando o outro interlocutor começa a falar. Esse recorte tem como vantagem utilizar uma delimitação que acontece naturalmente no curso das sessões clínicas e pode coincidir, muitas vezes, com a passagem de um conjunto de associações ligadas a um tema específico para outro tema similar, mas consideravelmente diferente. Sua desvantagem se deve à variabilidade no tamanho dos excertos que constituem um mesmo conjunto de associações, pois um turno de fala pode ser composto de uma ou duas palavras, no caso de uma concordância com o enunciado do interlocutor, por exemplo; ou pode ser extremamente longo e heterogêneo, como quando o analisando segue naturalmente fluxos associativos sem nenhuma intervenção por parte do analista.

O último predicado de caracterização exige que levemos em consideração as intervenções do analista durante a análise ao modelarmos o processo de associação livre. Como discutimos no Capítulo 1, a interpretação do analista tem por objetivo restituir a continuidade da experiência consciente, preenchendo as lacunas determina-



das pelo retorno do recalcado com as evidências de sua determinação inconsciente. Por se encontrarem na esteira das associações do analisando e por serem fonte de influência das associações subsequentes, as falas do analista precisam ser levadas em consideração, o que as torna, obrigatoriamente, parte do processo formal de livre associação.

### 3.1.2 Predicados de delimitação

Como um objeto modelo é necessariamente esquemático, sua classe de referência precisa delimitar propriedades relevantes e excluir outras, consideradas periféricas. Os predicados  $P_4$  e  $P_5$  servem para explicitar quais elementos da experiência da fala desconsideraremos no modelo. O primeiro dos predicados de delimitação nos desonera de levar em consideração a ordem na qual as palavras estão arranjadas em uma cadeia associativa e também nos permite desconsiderar as funções sintáticas exercidas pelos termos. Esse pressuposto implica que trabalharemos com um modelo linguístico do tipo *saco de palavras* (*bag of words*), pois o que nos interessa num contexto macro de associações – um turno de fala – é a frequência de ocorrência dos termos, e não sua ordem ou funções sintáticas.

As objeções a esses predicados precisam ser endereçadas. Primeiramente, é um pressuposto extremamente simplista acreditar que a ordem das palavras não carrega nenhuma informação que possa precisar a dimensão semântica dos termos. Apesar de a crítica ser relevante, precisamos deixar claro que o que nos interessa são os padrões globais de coocorrência dos termos, tais como descoberto pelo modelo de tópicos, e não a interpretação de uma passagem específica. De fato, se reduzirmos a análise das associações a um pequeno punhado de enunciados, a ordem das palavras poderá fazer uma grande diferença nas inferências possíveis. Mas do ponto de vista do conjunto total de associações disponíveis, o fato de um termo substantivo aparecer mais como objeto do que como sujeito nas orações não muda o fato de que ele tende a ocorrer junto de determinada classe de verbos, por exemplo. Outro argumento a favor de modelos mais simples se deve à natureza esparsa das informações linguísticas: frequências de coocorrência de termos já são bastante raras mesmo num *corpus* de tamanho considerável; quando consideramos a interação entre termos, sua ordem e função sintática, as relações existentes se repetem numa frequência ainda menor, tornando-se pouco informativa de padrões globais de associação. Temos

do nosso lado, também, o dado empírico de que modelo de saco de palavras tendem a ter um desempenho quase tão bom quanto modelos mais complexos em muitos problemas de processamento de linguagem natural (Sahlgren, 2006).

Por fim, a exclusão de aspectos prosódicos da fala e dos comportamentos não-verbais que ocorrem junto com ela é justificada para a simplificação do tipo de dado para o qual iremos propor nosso modelo teórico. Deixar de transcrever as nuances dos tons de voz e seu colorido emocional pode nos furtar de levar em consideração um material rico para a caracterização semântica dos termos, mas também nos poupa do grande trabalho extra que é anotar as transcrições com essas informações e nos permite simplificar consideravelmente a definição do modelo teórico. Afinal, há inúmeras maneiras possíveis de relacionar teoricamente as informações prosódicas com seu conteúdo verbal, e nenhum modelo específico da Psicanálise permite-nos reduzir as opções a ponto de torná-las implementáveis. Por outro lado, deixar de lado comportamentos associados à fala não é nenhuma surpresa para o trabalho psicanalítico. Mesmo que as ações do analisando possam ser de relevância na análise, é fazendo-as passar para o plano da fala que podemos interpretá-las propriamente. Ainda que algumas escolas psicanalíticas também se proponham a interpretar diretamente os atos do analisando, nos manteremos fiéis à tradição freudiana de focar a análise nos termos que compõem as associações.

### 3.1.3 Predicados de determinação

O último conjunto de predicados é necessário para justificar a elaboração de um modelo teórico que descreva o processo pelo qual as associações livres emergem. Sem esses predicados, seríamos obrigados a desenvolver uma perspectiva exclusivamente descritiva de nossos dados. Cada um desses predicados traz uma implicação importante para nosso objeto modelo, e por isso vamos abordá-los um a um.

O  $P_6$  postula a existência de um conjunto de entidades que *determinam* o processo de associação livre. Essas entidades são denominadas de *complexos associativos latentes*, tomando de empréstimo a noção de complexo utilizada por Freud e pela escola de Zurique para descrever o conjunto de ideias que influencia as respostas num teste de associação de palavras. Porém, indicamos com o predicado  $P_7$  que esses complexos são conjuntos de *termos* que

compartilham propriedades semânticas. Ainda que não tenhamos definido com clareza o que queremos dizer com “propriedades semânticas”, sua definição formal e operacional ficará clara quando desenvolvermos o modelo teórico, na próxima seção.

Os complexos associativos são necessariamente latentes porque não temos como observá-los diretamente: seja qual for a matéria que compõe o psiquismo – e não podemos deixar de lado o problema de que essa matéria precisaria ser bem descrita em termos neurológicos – simplesmente não existem meios para registrá-la, e só podemos inferir suas propriedades a partir de dados factuais, como as associações num processo de análise. Como partimos do pressuposto de que esses complexos se encontram efetivamente na origem das associações livres, não poderemos utilizar os resultados obtidos por nosso modelo para comprovar sua existência efetiva. Em última instância, convém lembrar, os complexos associativos são entidades postuladas em nosso objeto modelo para capturar as regularidades presentes nas associações livres, e de forma alguma precisam corresponder a uma entidade física de fato. Seu estatuto, portanto, é muito mais epistemológico – uma estratégia teórica para entender as coocorrências nas associações – do que propriamente ontológico.

A determinação proposta no predicado  $P_6$  deixa claro que o processo que faz emergir as associações dos complexos latentes não é puramente mecânico mas estocástico, ou seja, incorpora elementos aleatórios. Requerer um modelo determinístico das associações é uma exigência forte demais dada a quantidade de desconhecidos que simplesmente varreremos para debaixo do tapete na construção do modelo teórico. No lugar de tentar modelar todos os processos possíveis, postularemos alguns elementos centrais e deixaremos a variação restante por conta do acaso. Além disso, o pressuposto de que o processo associativo é estocástico permite-nos utilizar modelos probabilísticos para caracterizá-lo e a inferência bayesiana para estimar, a partir dos dados, os parâmetros desconhecidos do modelo.

Os últimos predicados,  $P_8$  e  $P_9$ , trazem algumas considerações qualitativas sobre o funcionamento dos complexos associativos. Primeiramente, estabelecemos a existência de uma hierarquia de importância entre os complexos, e que a ordem dessa hierarquia é dada por seu investimento libidinal. O termo *libido* é aqui utilizado de maneira mais flexível do que esperaríamos num contexto psicanalítico. Não nos interessa entrar no debate sobre a natureza sexual ou dessexualizada da libido, nem o problema ontológico de se ela é de fato uma energia ou não. O termo nos interessa, novamente,

em seu caráter epistemológico: na medida em que a experiência analítica mostra flutuações temporais na ordem de importância dos complexos, precisamos formalizar de alguma maneira essa maleabilidade. Ao mesmo tempo, é necessário deixar outros elementos do objeto modelo com maior fixidez, para que as regularidades detectadas sejam verossímeis. Por isso, postulamos que a ligação entre termos dos complexos associativos são consideravelmente estáveis e podem ser tratados como fixos; mas sua relevância, refletida em sua atividade na determinação no conteúdo das associações livres, pode mudar com o tempo – seja em função do processo de análise ou não. A variação temporal da relevância dos complexos associativos torna necessário registrarmos informações da ordem temporal do surgimento das associações, ou, mais precisamente, dos turnos de fala que as contêm.

### 3.1.4 Definição formal

Os predicados de caracterização e de delimitação permitem-nos representar os dados das associações livres de uma maneira tradicionalmente utilizada no processamento de linguagem natural. Como reduzimos as palavras a um conjunto de termos que compõe um vocabulário, caracterizamos unidades macro das associações nos turnos de fala e desconsideramos qualquer informações sobre a ordem das palavras ou elementos não-verbais, as associações livres podem ser representadas como uma *matriz de termos e documentos*. Nosso objeto modelo das associações livres passa a ser representado formalmente por uma matriz  $\mathbf{M}$  de dimensões  $V \times D$ , ou seja,  $V$  linhas, correspondentes a cada um dos termos presentes no vocabulário utilizado na análise, e  $D$  colunas, indicando cada um dos turnos de fala de todas as sessões em consideração. Cada elemento  $m_{ij}$  dessa matriz é um número inteiro e não-negativo, indicando quantas vezes o termo  $i$  aparece no turno de fala  $j$ .

Complementaremos a matriz de termos e documentos com uma matriz de metadados relevantes para os predicados de determinação, denominada  $\mathbf{X}$ . Os metadados são compostos, primeiramente, por uma variável binária indicando a qual dos interlocutores pertence um determinado turno de fala. Dessa maneira, podemos incluir na análise as falas do analista e avaliar o impacto com que as interpretações e outras intervenções podem ter no direcionamento das associações do analisando. Também utilizamos uma variável indicando à qual sessão pertence o turno de fala, de forma a permitir

avaliar a variação temporal das associações ao longo do tratamento. Por essas definições, a matriz  $\mathbf{X}$  terá  $D \times 2$  dimensões.

Nosso objeto modelo também pressupõe a existência de dois conjuntos de variáveis latentes. O primeiro conjunto diz respeito aos complexos associativos. Supondo haver um número finito  $K$  de complexos associativos ativos no processo de análise e assumindo que o vocabulário utilizado durante o tratamento também é fixo, com tamanho  $V$ , definimos a matriz  $\beta$ , de dimensões  $V \times K$ , em que cada coluna  $k$  é uma distribuição de probabilidade sobre os termos do vocabulário, de tal forma que termos representativos do complexo  $\beta_k$  tem alta probabilidade de ocorrerem em contextos similares.

O segundo conjunto de variáveis latentes se refere ao investimento libidinal sobre os complexos em cada turno de fala. Como já definido, partindo dos  $D$  turnos de fala e  $K$  complexos de um tratamento, definimos uma matriz  $\Theta$ , de dimensões  $K \times D$ , de tal forma que cada coluna  $d$  seja uma distribuição de probabilidade sobre os diferentes complexos. Complexos com maior probabilidade são considerados como tendo maior investimento libidinal e tendem a influenciar com mais força as associações livres, mas podem sofrer alterações de um turno ao outro ou, mais especificamente, de uma sessão para outra.

Nosso objeto modelo pode ser então formalmente representado por meio da seguinte expressão:

$$(\mathbf{M}, \mathbf{X}, \beta, \Theta) \triangleq R \quad (3.1)$$

Ou seja, o referente  $R$ , as associações livres que acontecem nas sessões de psicanálise, pode ser *modelado* pela tupla  $(\mathbf{M}, \mathbf{X}, \beta, \Theta)$ , em que cada elemento é uma das matrizes definidas acima, tendo em vista o conjunto de predicados  $P$  elencados e explicados anteriormente.

## 3.2 Modelo digital das associações livres

Precisamos, agora, inserir o objeto modelo elaborado para as associações livres dentro da teoria geral dos modelos de tópicos. É por meio dessa inserção que poderemos elaborar um *modelo digital*, uma descrição matemática e formal do processo que leva à ocorrência das associações livres. Pautados nas descrições feitas por Freud sobre os processos latentes que determinam as cadeias associativas, e baseados nos modelos de tópicos e em seus conceitos da área de

probabilidade, propomos um modelo digital do processo de associação livre. Esse modelo será inicialmente descrito de forma gerativa para facilitar sua elaboração. Para não complicar ainda mais a notação, utilizamos símbolos idênticos aos apresentados no capítulo anterior, aproximando os termos entre os dois modelos.

### 3.2.1 Dos complexos associativos

Na descrição do modelo têxtil do psiquismo, Freud concebe o conteúdo das diversas instâncias do aparelho psíquico como cadeias de traços organizadas em redes com múltiplos entrecruzamentos. Ao conjunto de traços próximos Freud dá o nome de complexo, indicando a existência de conjuntos de representações ligados por uma relação mais profunda do que aquela dada pela superfície linguística.

Em nosso objeto modelo, assumimos que as representações psíquicas que formam as cadeias e os complexos associativos são formadas por termos, e surgem no processo de livre associação exclusivamente sob a forma de palavras. Um complexo associativo é, então, um conjunto de termos  $V$  que podem ser facilmente substituídos entre si num mesmo contexto de associação. Em última instância, o complexo remete a um núcleo semântico que não é expresso diretamente, mas que pode ser identificado pelas diferentes palavras que ocorrem em contextos semelhantes durante as associações. Freud nos dá um exemplo disso quando fala de sua estratégia, já discutida na Seção 1.3, em pedir para o sonhador repetir o relato do sonho: ao relatar o sonho novamente, algumas substituições se tornam óbvias, pois diferem do relato original. Essas diferenças na superfície do relato revelam a estrutura semântica mais profunda que forma os complexos do sonhador e que permitem recuperar os pensamentos oníricos recalçados.

Em função dessa estrutura, podemos conceber que os complexos associativos se organizam no psiquismo como um número finito  $K$  de tópicos que reúnem termos semanticamente próximos. A proximidade semântica entre palavras se deve, justamente, à facilitação pela qual uma pode ser substituída por outra em contextos semelhantes. Se é a característica distribucional das palavras que permite deduzir suas relações semânticas, tal como pressuposto pelos modelos de tópicos, também é possível reconstituir os complexos inconscientes a partir das regularidades de substituições nas cadeias associativas. Os complexos podem ser modelados, portanto, como distribuições de probabilidade multinomiais  $\beta$  sobre um conjunto fi-

nito de termos  $V$ , com a probabilidade de cada termo nos complexos indicando o grau de facilidade com que um pode ser substituído por outro no processo de associação, sugerindo sua proximidade semântica. A proximidade semântica entre dois termos significa, portanto, que ambos os termos possuem uma alta probabilidade em um ou mais complexos associativos  $\beta$ , e poderiam surgir um no lugar do outro em um mesmo contexto de associação.

Como definimos no objeto modelo, os complexos estão dados *a priori* à intervenção do analista. Mesmo que por alguma influência ele possa interferir nas regularidades distributivas dos termos empregados pelo paciente, a estrutura já constituída da subjetividade sobrepujará qualquer sugestão inicial. Assumimos então que, ao longo de uma análise, os complexos são mais ou menos estáveis, e que podemos modelá-los como estáticos ao longo de todo o período considerado. O que muda, como veremos a seguir, é o investimento em cada um desses complexos ao longo das sessões. Isso significa, formalmente, que o parâmetro da distribuição de probabilidade multinomial  $\beta$  é fixo durante toda a análise.

Não entraremos no mérito de onde surgem e como se organizam geneticamente esses complexos. A Psicanálise possui uma teoria para compreender a constituição da subjetividade a partir da história das relações do sujeito, mas deixaremos esse aspecto de lado no modelo, pois com ele só podemos trabalhar tal qual ele se encontra em seu estado atual. Reconstruir sua gênese escapa do escopo de nossa proposta.

### 3.2.2 Das representações-meta

A partir do momento em que o psicanalista enuncia para o analisando a regra fundamental da Psicanálise, assumiremos, com Freud, que ele será levado a associar motivado por representações-meta inconscientes. Entenderemos as representações-meta, no escopo desse modelo, como o núcleo semântico que organiza cada um dos complexos associativos. Num determinado momento da vida do sujeito, diferentes complexos estarão mais investidos de energia psíquica. Quando o sujeito abandona suas intenções conscientes, a força do investimento desses complexos atrairá o processo de associação para si, respeitando os limites impostos pela censura.

Portanto, supomos que em cada sessão o analisando possui diferentes graus de investimento sobre seus complexos. Utilizaremos uma distribuição de probabilidade  $\theta$  para representar o *quantum* de

libido investido em cada um dos complexos. Assumiremos também que o investimento pode flutuar suave ou abruptamente. Por isso, ao contrário da distribuição  $\beta$ , que é fixa para todo um caso clínico,  $\theta$  será modelada para cada turno de fala, o que significa que teremos tantos  $\theta$  quanto turnos de fala  $D$  que estiverem sendo analisados, formando a matriz  $\Theta$ . Como nos interessa investigar como a proporção de tópicos abordados durante as sessões varia com o tempo, utilizaremos o número da sessão como covariável para prever essa proporção.

A probabilidade é uma aproximação possível para o investimento libidinal porque, assumindo que o total de libido é unitário, podemos interpretar cada quantidade investida como uma proporção dessa unidade. A utilização de uma distribuição de probabilidade sobre os complexos associativos também permite modelar a tendência do discurso do analisando em gravitar em torno de eixos semânticos principais. Em função da censura, não esperamos que um único tópico domine toda uma sessão, mas assumiremos que complexos candentes servem de representação-meta, atraindo as cadeias associativas para si, mesmo que desviando por outros complexos próximos.

Conceberemos também que as representações-meta poderão ser mais ou menos próximas e se influenciar mutuamente. Isso significa que um complexo não é completamente independente dos outros. Pelo contrário, alguns terão núcleos semânticos semelhantes e por isso poderão servir de facilitação no deslocamento no processo associativo. Esse pressuposto exige modelar os tópicos-complexos como correlacionados.

### 3.2.3 Das associações livres

Tendo abandonado suas representações-meta conscientes, o sujeito tem seu discurso determinado pelas representações-meta inconscientes dadas por seus complexos. O grau de investimento psíquico de um complexo influenciará o sujeito a produzir cadeias associativas ligadas a ele. A linguagem da probabilidade é particularmente útil para modelar esse processo: dada uma distribuição de probabilidade  $\theta$  sobre os complexos inconscientes – que representa o investimento psíquico num dado momento em cada um dos complexos – o sujeito tenderá a produzir associações estreitamente ligadas àqueles complexos com maior probabilidade.

Podemos modelar o processo de produção de associações li-



vres como uma cadeia de Markov cuja matriz de transição é dada por  $\theta$ . Mais especificamente, o processo de associação segue um modelo oculto de Markov (Andrews & Vigliocco, 2010), no qual a emissão de palavras depende tanto do complexo atuante no momento quanto do complexo que atuou na associação anterior. Numa dada sessão, ou seja, condicionado a um valor específico de  $\theta$ , cada termo  $w$  enunciado durante um turno de fala  $\vec{w}$  pode ser modelado como um passo da cadeia. Esse termo emerge, é claro, a partir dos termos que compõem os complexos do sujeito. A partir de um dado termo, a palavra seguinte terá seu complexo dominante determinado novamente por  $\theta$ . Com isso, podemos esperar que, mesmo quando algumas associações estão dominadas por complexos de menor relevância no momento, as associações tenderão a retornar aos complexos mais importantes. Utilizando o modelo oculto de Markov poderemos também identificar complexos relevantes que também replem as associações.

Esse pressuposto nos permite inferir do processo de associação livre a estrutura latente que desejamos. É justamente pelo fato de o sujeito ser atraído a produzir palavras ligadas aos complexos mais prementes que podemos deduzir quais desses complexos se encontram mais ativos durante um dado turno de fala, e, por consequência, numa sessão. Da mesma forma, tendo inferido a proporção dos complexos ativos num dado momento, podemos descobrir quais termos estão ligados a seu núcleo semântico.

### 3.2.4 Síntese do modelo digital

Retomando o modelo gerativo da Alocação de Dirichlet Latente, sintetizaremos nosso modelo para as associações livres, apresentando as ampliações que julgamos necessárias para melhor formalizar o processo de livres associações. O modelo resultante utiliza elementos do *Hidden Markov Topic Model* de Andrews e Vigliocco (2010) para dar conta das transições entre complexos dentro de uma mesma sessão, flexibilizando o pressuposto de saco de palavras e utilizando informações contextuais para melhor atribuir os tópicos. A utilização de covariáveis relativas ao tempo de tratamento para identificar efeitos temporais sobre a proporção de tópicos faz uso de alternativas à LDA propostas por Roberts et al. (2013) em seu *Structural Topic Model*.

1. Para cada complexo associativo  $k \in \{1, \dots, K\}$ ,

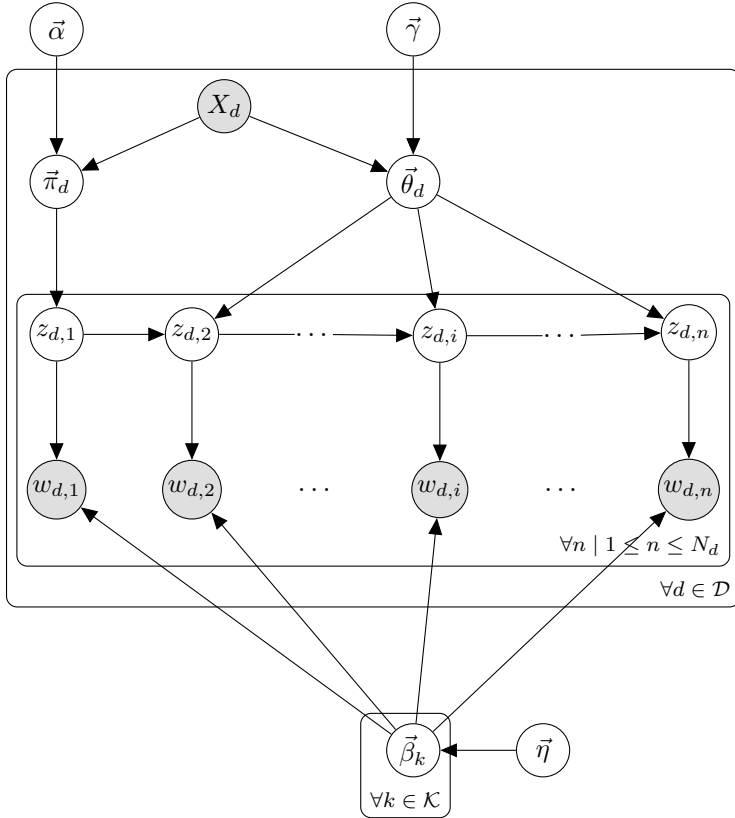
- a) Estabelecer uma distribuição sobre termos  $\vec{\beta}_k \sim \text{Dir}_V(\eta)$ .
- 2. Para cada turno de fala  $\vec{w}_d \in \mathcal{D}$ , com covariável temporal  $X_d$ ,
  - a) Estabelecer um vetor de investimentos sobre os complexos para o estado inicial das associações,  $\vec{\pi}_d \sim \text{Dir}_K(X_d \alpha)$ .
  - b) Estabelecer uma matriz de transição de estado de investimentos sobre os complexos considerando a covariável temporal  $X$ ,  $\vec{\theta}_d \sim \text{Dir}_K(X_d \gamma)$ .
  - c) Para cada associação  $w_{d,n} \in \vec{w}_d$ ,
    - i. Escolher uma representação-meta  $z_{d,n}$  condicionada à representação-meta anterior,  $z_{d,n-1}$ ,  $z_{d,n} \sim \text{Mult}(\vec{\theta}_d | z_{d,n-1})$ ,  $z_{d,n} \in \{1, \dots, K\}$ .
    - ii. Associar uma palavra  $w_{d,n} \sim \text{Mult}(\vec{\beta}_{z_{d,n}})$ ,  $w_{d,n} \in \{1, \dots, V\}$ .

### 3.2.4.1 Modelo gráfico

Para ilustrar melhor a relação entre variáveis e diferenciar nosso modelo da LDA tradicional, apresentamos na Figura 3.1 o modelo gráfico probabilístico que descreve o modelo gerativo sintetizado anteriormente.

## 3.3 Implementação do modelo: o pacote `stm`

A implementação do modelo proposto, contudo, é particularmente complicada em função da necessidade do cálculo da matriz de transição do modelo oculto de Markov proposto para as associações livres. Em seu lugar, aproximaremos o modelo proposto pelo *Modelo de Tópicos Estruturais* (*Structural Topic Model*, STM), brevemente apresentados na Subseção 2.3.7, bastante semelhante mas relativamente mais simples, e já implementado no pacote `stm`, disponível para a linguagem de programação R (Roberts, Stewart, & Airolidi, 2016). O STM foi desenvolvido para capturar diretamente o efeito de covariáveis e metadados sobre a distribuição dos termos nos tópicos e tópicos nos documentos. As covariáveis podem afetar tanto a *prevalência dos tópicos* em cada documento quanto o *conteúdo dos*



**Figura 3.1:** Modelo de tópicos para as associações livres

*tópicos*, ou seja, quais termos estão mais fortemente associados a cada tópico.

Ampliando o modelo da LDA, apresentado em detalhes no capítulo anterior, o STM pode ser dividido em três módulos: um *modelo da prevalência* de tópicos, baseado na regressão Dirichlet-Multinomial e determinando como tópicos são alocados em documentos; um *modelo do conteúdo* de tópicos, baseado no Modelo Aditivo Geral Esparsa, que estrutura a maneira como termos são atribuídos aos tópicos; e um *modelo de linguagem*, que explicita como os tópicos produzidos pelo modelo de conteúdo e a proporção de tópicos produzida pelo modelo de prevalência são instanciados em documen-

tos específicos a partir das palavras do vocabulário (Roberts et al., 2013).

O modelo de prevalência de tópicos substitui a distribuição de Dirichlet que rege a mistura de tópicos em cada documento por uma regressão Dirichlet-Multinomial. Esse modelo torna a distribuição *a priori* sobre a proporção de tópicos dependente de uma função linear de um número arbitrário de covariáveis. O STM expande o modelo Dirichlet-Multinomial utilizando uma distribuição normal-logística cujos valores esperados são o resultado das equações dos modelos lineares generalizados, podendo fazer uso de funções de base não-lineares, como *splines* cúbicas; e cuja matriz de covariância permite ajustar termos de correlação entre os tópicos. Dessa maneira, trocamos a matriz de transição do modelo oculto de Markov por uma matriz de covariância entre tópicos, tornando possível avaliar a dinâmica de coocorrência dos complexos associativos.

A distribuição de Dirichlet que determina a alocação dos termos nos tópicos, por sua vez, é substituída pelo Modelo Aditivo Geral Esparso, permitindo o conteúdo dos tópicos variar a partir de covariáveis. Esse modelo constrói uma linha de base da distribuição dos termos do vocabulário, a partir da qual cada tópico, em função das covariáveis, é um desvio aditivo esparsos, regularizado, na escala logarítmica. A regularização evita que palavras raras tenham um peso muito grande em um tópico específico, e facilita a adição de um número arbitrário de preditores. Para nossos fins, porém, não utilizaremos esse módulo do modelo, pois assumimos que os complexos associativos são fixos e não variam em função de nenhuma covariável de interesse. Por isso, manteremos o modelo de tópicos baseados na distribuição de Dirichlet.

O modelo de linguagem do STM é baseado na distribuição normal-logística para modelar a proporção de tópicos em cada documento. Em um modelo com  $K$  tópicos, um vetor  $\vec{\eta}_d$  com  $K - 1$  dimensões, é retirado de uma distribuição normal para cada documento do *corpus*,  $\vec{\eta}_d \sim N(\vec{\mu}_d, \Sigma)$ . O vetor  $\vec{\mu}_d$  é o valor esperado da transformação logit da proporção de cada tópico para aquele documento, em função de suas covariáveis, dado pela equação  $\vec{\mu}_d = \vec{\Gamma}' \vec{x}_d$  – o produto matricial do vetor de coeficientes de regressão  $\vec{\Gamma}$ , comum a todos os documentos, e o vetor de covariáveis  $\vec{x}$ , específico ao documento  $d$ .  $\Sigma$  é uma matriz quadrada de dimensões  $(K - 1) \times (K - 1)$ , e indica a covariância entre os tópicos do modelo, o que permite estimar sua correlação. O STM aplica uma regularização sobre essa matriz para que ela seja esparsa, evitando assim correlações espú-

rias entre tópicos. A força da regularização pode ser controlada no ajuste do modelo pela definição dos valores do hiperparâmetro  $\rho$ .

O uso do modelo normal logístico exige trabalhar com uma dimensão a menos do que o total,  $K - 1$ , para permitir a identificação do modelo, pois a soma das proporções precisa, necessariamente, ser igual a 1. O vetor  $\vec{\eta}_d$  é mapeado para o simplex pela transformação logística:  $\theta_{d,k} = \exp(\eta_{d,k}) / \sum_{i=1}^K \exp(\eta_{d,i})$ . Como o valor  $\eta_{d,K}$  é fixo em 0, pelo motivo apontado anteriormente, o vetor  $\vec{\theta}_d$  irá, necessariamente, estar inscrito no simplex  $K - 1$ , e sua soma será igual a 1, como necessário para o parâmetro de uma distribuição multinomial.

O vetor de coeficientes da regressão Dirichlet-multinomial é regularizado, de forma a evitar o superajuste do modelo. Apesar da função utilizada para prever a expectativa de proporção de tópicos ser linear, o STM tem suporte para transformação não-lineares, como *B-splines*, de forma a poder modelar relações complexas entre covariáveis de proporção e tópicos. A inclusão de covariáveis também ajuda a tornar a estimação mais robusta em função do agregamento parcial, na medida em que documento com covariáveis similares ajudam a estimar os parâmetros de proporção de tópicos uns dos outros.

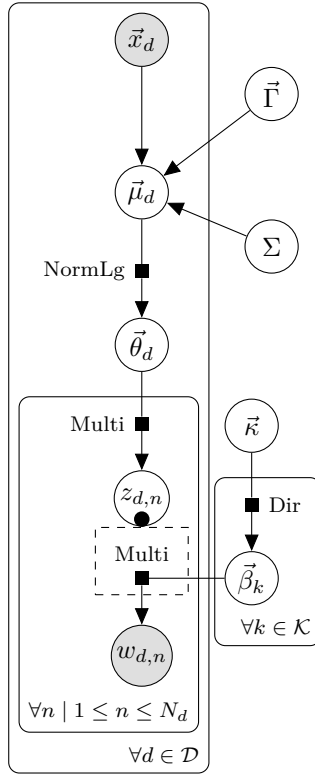
A distribuição *a priori* sobre os coeficientes da regressão linear para a proporção de tópicos é uma distribuição normal centrada em 0 e com variância compartilhada entre os diversos coeficientes do modelo,  $\gamma_{p,k} \sim N(0, \sigma_k^2)$ . Esses parâmetros de variância são diferentes para cada tópico  $k$  ao qual o coeficiente pertence, e é modelada a partir de uma distribuição gama inversa,  $\sigma_k^2 \sim \text{InvGama}(a, b)$ , em que os hiperparâmetros  $a$  e  $b$  podem ser definidos de forma a permitir maior variação dos coeficientes entre tópicos.

Uma vez amostrado o parâmetro  $\vec{\theta}_d$ , o modelo de linguagem do STM segue o mesmo caminho da LDA: um vetor de tópicos latentes  $\vec{z}_d$ , de dimensão igual ao número de palavras no documento, é sorteado a partir de uma distribuição multinomial,  $\vec{z}_d \sim \text{Multinomial}(\vec{\theta}_d)$ ; em seguida, um termo  $w_{d,n}$  é sorteado para cada elemento de  $\vec{z}_d$  utilizando a distribuição multinomial de tópicos  $\vec{\beta}_{z_{d,n}}$ .

A estimação dos parâmetro do modelo é feita por inferência bayesiana. A distribuição posterior dos parâmetros de interesse é definida pela seguinte proporcionalidade, se desconsiderarmos o modelo de conteúdo dos tópicos (adaptado de Roberts, Stewart, & Airoldi, 2016):

$$\begin{aligned}
p(\vec{\eta}, \vec{z}, \vec{\beta}, \Gamma, \Sigma \mid \vec{w}, \mathbf{X}) \propto \\
p(\Sigma) \prod p(\vec{\beta}) \prod p(\Gamma) \times \\
\left( \prod_{d=1}^D N(\vec{\eta}_d \mid \Gamma' \vec{x}, \Sigma) \left( \prod_{n=1}^N \text{Mult}(z_{d,n} \mid \vec{\theta}_d) \text{Mult}(w_n \mid \vec{\beta}_{d,k=z_{d,n}}) \right) \right)
\end{aligned} \tag{3.2}$$

Ou, alternativamente, como modelo gráfico apresentado na Figura 3.2.



**Figura 3.2:** Modelo gráfico probabilístico do STM baseado em Roberts et al. (2013)

Como já vimos com relação à LDA, a resolução analítica da distribuição posterior é intratável, principalmente porque a distribuição normal-logística não é conjugada da distribuição multinomial. A estimativa dos parâmetros precisa ser feita de modo aproximativo, e o STM utiliza maximização-expectância variacional. A expectância é estimada por meio de aproximação de Laplace para os termos de proporção amostrados da distribuição logística normal. Para aumentar a velocidade do cálculo da expectância dos parâmetros da proporção de tópicos,  $\tilde{\eta}_d$ , o algoritmo marginaliza os vetores  $\tilde{z}_d$ , mas o reintroduz na etapa de maximização para computar a distribuição do tópicos  $\tilde{\beta}$ .

A aproximação variacional é dada pela distribuição  $\prod_d q(\tilde{\eta}_d)q(\tilde{z}_d)$ , em que o primeiro termo é modelado por uma distribuição normal, refletindo a equação do modelo original, e o segundo termo é modelado como uma distribuição multinomial. Na distribuição variacional, porém, os termos são independentes, facilitando o cálculo da expectância, mas exigindo aproximações no caso do parâmetro de proporções de tópicos.

Testes comparativos do STM com a LDA a partir de dados simulados, com a influência de uma covariável, indicam claramente a vantagem do STM recuperar estimativas mais precisas da proporção de tópico ao levar em consideração os efeitos da covariável. Outro estudo baseado em simulação, mas a partir das estimativas obtidas da aplicação do modelo sobre dados reais, indicam que as estimativas produzidas pelo algoritmo de EM variacional produz intervalos de credibilidade com percentual de cobertura próximo ao seu valor nominal, indicando uma boa propriedade frequentista dessas estimativas. Aplicando o STM sobre dados reais e avaliando sua performance por meio de verossimilhança preditiva de excertos de documentos não utilizados no treinamento, o desempenho observado foi superior a outros modelos de tópicos, mesmo aqueles que também faziam uso de covariáveis para prever a proporção de tópicos e alocação de termos (Roberts, Stewart, & Airoldi, 2016).

### 3.3.1 Diferenças com o modelo proposto

A grande vantagem em utilizar o STM no lugar do modelo que propusemos acima é que os algoritmos para ajuste do modelo já estão implementados e testados no pacote `stm`, poupando-nos do trabalho de elaborar e testar um algoritmo próprio. A necessidade de utilizar aproximações variacionais para implementar nosso modelo

torna a tarefa praticamente impossível, por envolver a necessidade da computação de distribuições variacionais, uma habilidade que infelizmente está além de nosso alcance no momento. Mesmo assim, o STM é flexível o suficiente para aproximar-se muito bem de nossa proposta original. Precisamos, contudo, estar atentos a algumas mudanças cruciais para compreendermos melhor como estamos efetivamente modelando as associações livres.

A maior diferença é que o STM mantém o pressuposto de *bag of words*, exatamente como a LDA, e, portanto, descarta qualquer informação com relação à ordem das palavras nos documentos. Em função disso, não será possível estimar as probabilidades de transição entre um complexo e outro, como originalmente pretendido pela utilização de um Modelo Oculto de Markov para representar a dependência temporal entre tópicos latentes. Em seu lugar, aproveitaremos o fato de que o STM utiliza uma distribuição normal-logística como distribuição *a priori* para a proporção de tópicos nos documentos – ou, em nossos termos, um vetor de investimentos sobre complexos – e investigaremos sua matriz de covariância para identificar relações entre complexos: quais tendem a ocorrer em contextos semelhantes, indicando proximidade; quais possuem correlação negativa e, portanto, a dominância de um implica na evitação do outro. Em suma, ao utilizarmos o STM estaremos trocando as matrizes de transição entre tópicos por uma única matriz de covariância entre tópicos. Como o ajuste dessa matriz de covariância é particularmente sensível a superajustes, utilizaremos uma distribuição *a priori* para regularizar o resultado final em direção a uma matriz diagonal.

Essa pequena perda com relação ao modelo original é compensada por uma outra diferença importante: no lugar de utilizarmos apenas uma variável temporal para indicar a ordem de ocorrência das sessões,  $X_d$ , utilizaremos a implementação da regressão Dirichlet-Multinomial para acrescentarmos mais metadados ao modelo final. Utilizaremos duas informações cruciais: (1) o número da sessão à qual pertence o turno de fala; e (2) o interlocutor ao qual pertence o turno de fala. Desse modo, forçamos no modelo uma forte dependência entre turnos de fala de uma mesma sessão – o que é desejável, pois não faz sentido acreditar que a distribuição de investimentos sobre os complexos se alteraria bruscamente dentro de uma única sessão. Incluindo informação sobre o interlocutor que enunciou o turno de fala, poderemos separar melhor os tópicos que surgem com pouca influência do analista, aqueles que pertencem ex-



clusiva ou quase que exclusivamente às interpretações do analista, e aqueles tópicos que são prováveis tanto na fala do analisando quanto do analista, o que pode indicar uma possível contaminação das associações por parte das intervenções. A regressão Dirichlet-Multinomial – ou, mais corretamente, Normal-Logística-Multinomial – que utilizaremos no ajuste do modelo é definida na Equação 3.3.

$$\vec{\mu}_d \mid x_d^{inter}, x_d^{sess\tilde{a}o} = \vec{\gamma}^{inter} x_d^{inter} + \vec{\gamma}^{sess\tilde{a}o} S(x_d^{inter}) \quad (3.3)$$

A ideia do modelo de regressão para a proporção de tópicos em cada documento é simples: o valor esperado da distribuição de tópicos em um documento  $d$ ,  $\vec{\mu}_d$ , dado um identificador do interlocutor,  $x_d^{inter}$ , e o número da sessão à qual pertence  $d$ ,  $x_d^{sess\tilde{a}o}$ , é igual a uma constante para cada interlocutor,  $\vec{\gamma}^{inter} x_d^{inter}$  – elementos do vetor  $\vec{\gamma}^{inter}$  que forem positivos indicarão tópicos mais utilizados pelo analista, elementos negativos indicarão tópicos mais utilizados pelo analisando e elementos próximos a zero indicarão tópicos utilizados por ambos – somada aos coeficientes relativos à sessão,  $\vec{\gamma}^{sess\tilde{a}o} S(x_d^{inter})$ , utilizando uma *b-spline* para lidar com variações não lineares dos tópicos ao longo do tempo. Como temos apenas 47 sessões para identificar a passagem do tempo, utilizamos 10 graus de liberdade para a *b-spline*, utilizando a base polinomial cúbica por padrão.

### 3.4 Procedimentos do estudo de caso

A avaliação do modelo de tópicos para as associações livres só pode ser realizada por meio do ajuste do modelo a dados provenientes de um estudo de caso. O modelo proposto é útil em organizar e estabelecer limites para as relações entre os elementos que nos interessam – principalmente entre complexos associativos e as cadeias associativas produzidas durante uma psicanálise – mas é geral demais para permitir inferências precisas. Essas inferências se tornam possíveis, porém, se instanciarmos o modelo por meio de um estudo de caso, limitando o escopo das inferências possíveis mas enriquecendo o modelo com associações efetivas enunciadas durante um processo de análise.

### 3.4.1 Protocolo e contato inicial

Para o protocolo inicial do estudo de caso, decidimos pelo registro em áudio de 50 sessões como um número ideal para equilibrar o custo da transcrição das entrevistas e a obtenção de uma quantidade suficiente de dados para ajustar o modelo. O número de sessões necessárias para a pesquisa da tese não foi colocado como limite de duração do tratamento.

Frente à dificuldade de conseguir um psicanalista experiente que aceitasse gravar suas sessões com um paciente, decidimos que o próprio autor da tese, psicanalista de orientação freudo-lacanianiana com cerca de dez anos de experiência clínica, conduziria uma análise voltada para a obtenção de dados para a pesquisa. O protocolo também previa a realização de testes de associação de palavra com o paciente a partir do momento em que pelo menos 30% das sessões estivessem transcritas.

Com relação ao analisando, o protocolo inicial impunha apenas que fosse adulto, buscasse o tratamento por vontade própria e não estivesse tomando nenhum tipo de medicamento psiquiátrico. As entrevistas preliminares deveriam servir também para garantir não se tratar de um paciente psicótico, o que exigiria a procura por um novo caso.

A pesquisa foi realizada em um serviço de atendimento psicológico vinculado a um curso de graduação em Psicologia. O serviço atende gratuitamente membros da comunidade acadêmica e da comunidade externa à universidade. Por se tratar de uma clínica-escola, os potenciais participantes foram os usuários na fila de espera, que foram contatados por telefone e esclarecidos sobre o objetivo da pesquisa antes do agendamento da entrevista inicial.

O analisando cujo caso é apresentado nesta tese foi o segundo usuário contatado, após o primeiro usuário da fila de espera afirmar não estar mais interessado em realizar um tratamento psicoterápico. Esclarecido sobre a utilização dos dados de seu tratamento para a realização da presente pesquisa, o paciente se mostrou disposto a participar desde o primeiro contato telefônico. Apesar do uso dos dados para pesquisa, os atendimentos foram conduzidos como qualquer outra psicanálise, com a possível exceção do uso do recurso do gravador.

Próximo ao término do ano, o analisando decidiu retornar à sua cidade natal e, por escolha própria, encerrou os atendimentos ao fim da 47<sup>a</sup> sessão, tendo comunicado antecipadamente ao analista o tempo que ainda permaneceria sob tratamento. Por isso, o estudo

de caso que serve de prova de conceito da aplicação de nosso modelo digital das associações livres é baseado em 47 sessões da psicanálise do sujeito de pesquisa, um jovem obsessivo, conduzidas de maio a novembro de 2015. As sessões tiveram duração variando de 20 a 55 minutos, e foram realizadas de duas a três vezes por semana. Detalhes descritivos das sessões são apresentados no próximo capítulo, junto com uma descrição exploratória do *corpus* produzido a partir dos atendimentos. Informações específicas sobre o analisando são apresentados no Capítulo 5.

### 3.4.2 Condução das sessões

Seguindo a recomendação dada por Freud (1984h), as entrevistas preliminares e as sessões foram conduzidas com nosso sujeito de pesquisa sempre dentro da regra fundamental da Psicanálise. A ênfase na regra fundamental foi a mais radical possível, reduzindo ao máximo as intervenções do psicanalista e enfatizando a direção do tratamento pelas associações livres do analisando, numa estratégia mais próxima da técnica de análise das resistências (Freud, 1984f). Ainda que se possa objetar que as entrevistas preliminares não implicam que o sujeito está efetivamente em análise, não utilizamos aqui essa distinção. Afinal de contas, se conduzimos o tratamento conforme a regra fundamental, as associações produzidas já são informativas dos complexos inconscientes e das resistências que agem sobre o sujeito.

O objeto modelo proposto e sua descrição como modelo digital aproximada pelo STM fazem referência às associações livres do produzidas ao longo de uma análise. Entretanto, não temos como analisar diretamente as associações produzidas durante as sessões. Em função disso, precisamos substituí-las por algo que conserve alguma proximidade aceitável com o que foi dito. A solução mais imediata foi a transcrição das gravações das sessões de um caso clínico. Objeções a esse procedimento são usuais no contexto da pesquisa em Psicanálise (Silva et al., 2014), mas caem por terra se considerarmos os resultados das investigações de Thomä e Kächele (Dahl et al., 1988), que não conseguiram detectar nenhum efeito intrusivo significativo da gravação sobre a análise. Apesar das transcrições perderem informações relativas a uma série de aspectos relevantes da linguagem, tais como a entonação, os gaguejos e as interrupções, ainda assim conseguimos preservar um aspecto crucial: a distribuição de coocorrência entre termos, que é o que interessa para o ajuste

do modelo. Com o objetivo de transpôr as entrevistas em um *corpus* textual, portanto, gravamos cada sessão integralmente, com a anuência do analisando.

### 3.4.3 Transcrição

Todas as entrevistas preliminares e sessões foram transcritas integralmente, com auxílio do *software* Audacity, um editor de áudio que pode ser utilizado para otimizar o processo de transcrição. Foram registrados os turnos de fala do analista e do analisando, devidamente identificados para processamento posterior. Maneirismos ligados ao uso oral da língua foram todos vertidos para a forma ortográfica, de forma a facilitar a padronização das palavras em termos. Como o modelo proposto está focado na distribuição das palavras, alguns elementos não-verbais do discurso foram denotados por marcadores padronizados entre colchetes. Silêncios e interrupções superiores a dois segundos foram registrados por reticências ([. . .]); risos, pigarros e outros sons também foram registrados com a marcação de colchetes (p.e., [risos]). Outros elementos, como entonação e maneirismos, mais difíceis de transcrever em conjunto com as palavras, não foram considerados para a composição do *corpus*.

Junto com as palavras transcritas foram anotados metadados das sessões: data e hora, número da sessão e tempo de duração. Alguns metadados da sessão, em conjunto com alguns metadados dos turnos de fala, foram utilizados para composição de matriz de preditores da equação da proporção de tópicos em cada documento, conforme explicitado na Subseção 3.1.4.

Foram tomados os devidos cuidados para substituir elementos que possam identificar o analisando. Nomes próprios, nomes de lugares e outros indicativos pessoais foram substituídos por dados fictícios.

### 3.4.4 Teste de associação de palavras

Um dos principais pontos questionáveis do modelo é se os tópicos encontrados podem ser considerados como correspondentes aos complexos no psiquismo do sujeito. Na ausência de um critério bem delimitado para a validação de modelos de tópicos no contexto das psicoterapias, apoiamo-nos na referência histórica e crucial, para a Psicanálise, do estabelecimento da relação entre os complexos inconscientes e as associações livres como endereçado pela aplica-

ção do teste de associação de palavras por Jung, cujos resultados foram tomados por Freud como comprovação experimental da técnica de associação livre, conforme exposto em maiores detalhes na Subseção 1.3.3.

Procedemos de maneira semelhante: para colocar à prova o valor da evidência das informações obtidas pela aplicação do modelo de tópicos, utilizamos resultados de testes de associação de palavras, que não foram acrescentados ao *corpus* ou aos metadados. Foram conduzidos cinco testes de associações de palavras com o analisando, sempre 10 minutos antes de cada uma das cinco últimas sessões. Utilizando os dados das 20 sessões transcritas até aquele momento, foram selecionadas aleatoriamente 20 palavras-estímulos para cada teste dentre todas as palavras com grande frequência de ocorrência nas associações do analisando, excluindo termos com pouco conteúdo semântico, como conjunções e verbos auxiliares. Como a escolha foi aleatória, algumas palavras aparecem em mais de uma aplicação do teste. O analisando foi esclarecido sobre o objetivo do teste e foi instruído a responder com a primeira palavra que lhe viesse à cabeça, assim que o analista terminasse de ler a palavra-estímulo. O tempo de reação do analisando também foi registrado. Foram obtidos, portanto, 100 pares de palavras-estímulo e palavras-resposta.

Os resultados dos testes de associação de palavras e suas consequências para a validade do modelo de tópicos são apresentados e discutidos no próximo capítulo.

### 3.4.5 Local de realização da pesquisa e considerações éticas

A abordagem do sujeito de pesquisa seguiu as orientações preconizadas pela CONEP para pesquisa com seres humanos. O processo de consentimento livre e esclarecido foi iniciado desde o primeiro contato telefônico com o sujeito na fila de espera do serviço de atendimento. Antes mesmo da primeira sessão, o sujeito foi informado de que os dados seriam utilizados para uma pesquisa acadêmica, sendo sua identidade e anonimato resguardados, e só foram marcadas as entrevistas preliminares quando o sujeito se mostrou interessado em participar. Na primeira entrevista, a natureza da pesquisa foi melhor explicada, bem como seus benefícios e possíveis problemas. O sujeito foi informado de que as sessões seriam gravadas em áudio, e aceitou participar mesmo sob essas condições.



## Capítulo 4

# Seleção e validação do modelo

Antes de iniciarmos as explorações do *corpus* por meio do modelo de tópicos para as associações livres, é necessário apresentar algumas características descritivas gerais dos dados obtidos. Em seguida, abordamos os procedimentos utilizados para tratar os dados brutos, no sentido de remover termos de pouco valor semântico (*stopwords*), identificar pares de palavras que formam locuções (*collocations*) e padronizar os termos em sua forma canônica (lematização). O tratamento inicial é crucial para o bom funcionamento do algoritmo de estimação do modelo.

Porém, apenas o tratamento dos dados brutos não é suficiente. Como os modelos de tópicos são sensíveis ao *número de tópicos* inicialmente definido e à *multimodalidade* da distribuição posterior, avaliamos o ajuste de diversos modelos semelhantes para identificar o número de tópicos mais adequado ao *corpus* analisado e também qual das soluções encontradas apresenta características desejáveis. Para não incorrer numa avaliação subjetiva do que seria um aspecto desejável do modelo ajustado, os critérios de seleção são definidos formalmente e a seleção é feita sem a intervenção do pesquisador quanto à interpretabilidade dos resultados.

Por fim, o modelo final precisa ser avaliado a partir de informações externas a ele. Como os modelos de tópicos foram inicialmente concebidos como ferramenta exploratória voltada para a aprendizagem não-supervisionada de dados com pouca estrutura, a validação externa é condição necessária para que possamos ter alguma confiança nos resultados. Ao mesmo tempo, como não há uma literatura bem desenvolvida sobre como validar modelos de tópicos na exploração de transcrições de sessões clínicas, é também a tarefa mais desafiadora desta pesquisa, pois exige conjugar o conhecimento específico da área com derivações de predições a partir do modelo formal. A avaliação é feita por meio da comparação entre as predições feitas pelo modelo ajustado e uma bateria de testes de

associações de palavras aplicada com o analisando.

## 4.1 Descrição do *corpus*

O *corpus* do estudo de caso é composto pela transcrição integral de 47 sessões de psicanálise, totalizando 31 horas e 12 minutos de áudio. A duração das sessões variaram de 20 a 55 minutos, com média de 40 minutos. Foram registrados 3.219 turnos de falas, 1.611 do analisando e 1.608 do analista. A presença de mais turnos de falas por parte do analisando se deve ao fato de que os primeiros turnos de fala sempre são do analisando, mas os turnos de fala finais, apesar de pertencerem, em sua maioria, ao analista, podem também pertencer ao analisando. O número de turnos de fala nas sessões variaram de 26 (na última sessão, de número 47) a 116 (na sessão 17), com média de 68 turnos por sessão (desvio-padrão de 21 turnos).

Para continuar com as análises, efetuamos apenas a eliminação dos sinais de pontuação e convertemos todas as palavras para minúsculas. O vocabulário do *corpus* bruto é composto por 9.768 termos, totalizando 221.187 palavras. 4.391 termos (45% do total) são *hapaxes*, ou seja, termos que aparecem uma única vez em todo o *corpus*. O número de palavras por turno varia de 1, tanto para analista quanto para analisando – correspondendo a turnos de fala curtos, expressando concordância, discordância ou questões monossilábicas – a 1.210 e 504 para analisando e analista, respectivamente. Em média, porém, os turnos de fala são compostos por 100 (DP=127) palavras para o analisando e 37 (DP=49) para o analista, indicando a tendência esperada do analisando fazer uso de turnos mais longos, mas com considerável variação no número de palavras. Um resumo dessas informações descritivas pode ser conferido na Tabela 4.1.

É interessante avaliar visualmente a contagem de palavras em cada turno de fala ao longo das sessões. Como pode ser conferido na Figura 4.1, há uma forte variação no número de palavras dentro de uma mesma sessão, como indicado pelos picos e vales presentes na série temporal, plotada na escala logarítmica para facilitar a visualização. Os números na parte superior dos gráficos indicam o número da sessão, plotados de cinco em cinco para facilitar a visualização. A linha em azul indica a tendência central não-linear ao longo do tempo. Com relação ao analista, essa tendência sofre um leve aumento no começo do tratamento, até a sessão 10, e fica estável até o término do tratamento. No caso do analisando, há uma considerável



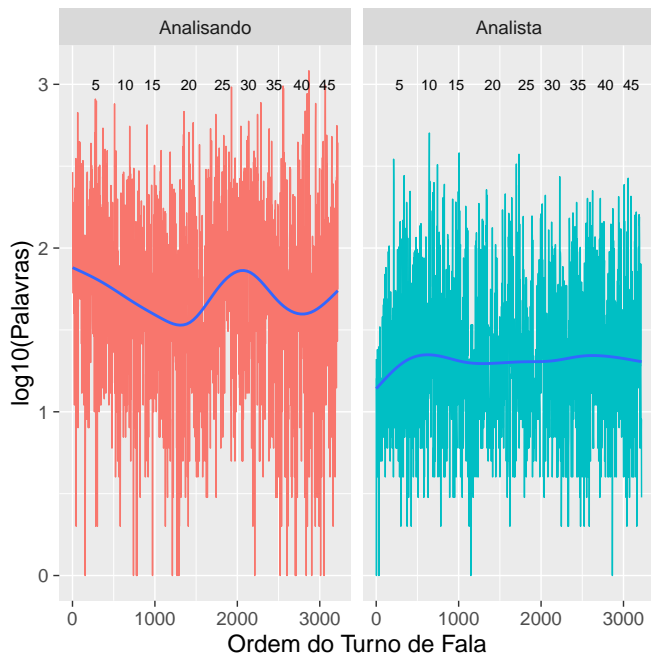
**Tabela 4.1:** Informações descritivas do *corpus* não tratado.

	Analizando	Analista	Total
Turnos de Fala	1.611	1.608	3.219
Média de TdF por Sessão	34	34	68
Mediana de TdF por Sessão	35	34	69
Vocabulário	6.817	5.740	9.768
Palavras	161.300	59.887	221.187
Média de Palavras por TdF	100	37	69
Mediana de Palavras por TdF	56	17	31

redução no número médio de palavras utilizadas em cada turno de fala até a sessão 20, aumentando e diminuindo rapidamente entre as sessões 20 e 40, e voltando a aumentar depois disso. Um possível motivo dessa flutuação será dado quando explorarmos a variação da força dos complexos ao longo do tratamento no próximo capítulo.

Uma outra alternativa para visualizar a distribuição do tamanho de cada turno de fala é por meio de um gráfico da função de probabilidade cumulativa empírica, apresentado na Figura 4.2. Nesse gráfico, o número de palavras por turno de fala foi transformado por meio do logaritmo de base 10, facilitando a identificação da proporção do número de documentos com um determinado tamanho. A linha azul indica os turnos de fala do analista e a linha vermelha indica os turnos de fala do analisando. A distribuição de palavras por turnos de fala é bastante regular no caso do analisando. Como é possível visualizar no gráfico, cerca de 12% dos turnos de fala do analisando possuem menos de 10 palavras, mas metade do turnos possuem de 30 a 100 palavras, de acordo com o intervalo interquartilico. No caso do analista, essa distribuição é bem menos regular e com número de palavras consideravelmente menor. Cerca de 25% dos turnos de falas do analista possuem menos de 10 palavras, com metade dos turnos variando de 10 a 50 palavras, também com base no intervalo interquartilico.

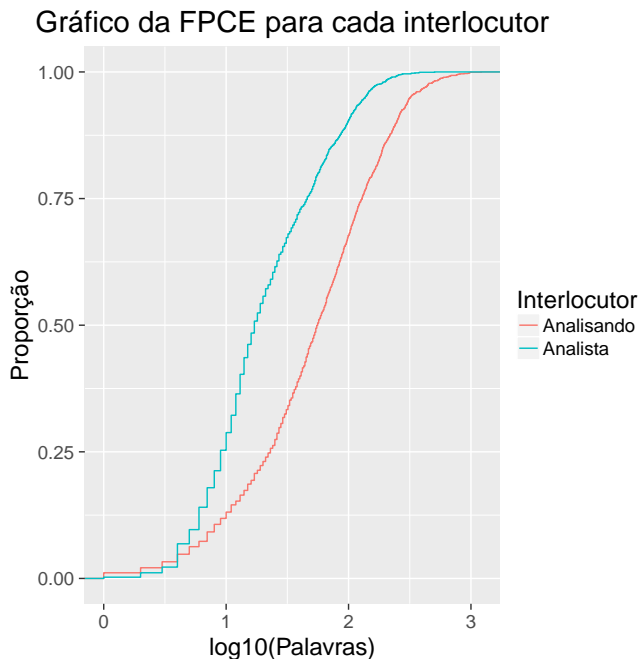
A visualização da distribuição das palavras, por sua vez, pode ser conferido por meio de um gráfico de Zipf, como na Figura 4.3. O gráfico de Zipf, popularizado pelo linguista norte-americano George Kingsley Zipf, representa o logaritmo na base 10 da contagem de frequência dos termos no eixo *y* e o logaritmo na base 10 do ranque dos termos no eixo *x*. Para a maioria das linguagens naturais, o gráfico de Zipf exibe um comportamento linear e inverso, indicando



**Figura 4.1:** Gráfico de linha do logaritmo da contagem de palavras em cada turno de fala do *corpus* não tratado. Sobre cada gráfico foi sobreposto uma linha de regressão não-linear para facilitar a visualização das tendências temporais. Os números na parte superior dos gráficos indicam as sessões.

que a frequência de uma palavra é inversamente proporcional a sua ordem no ranque.

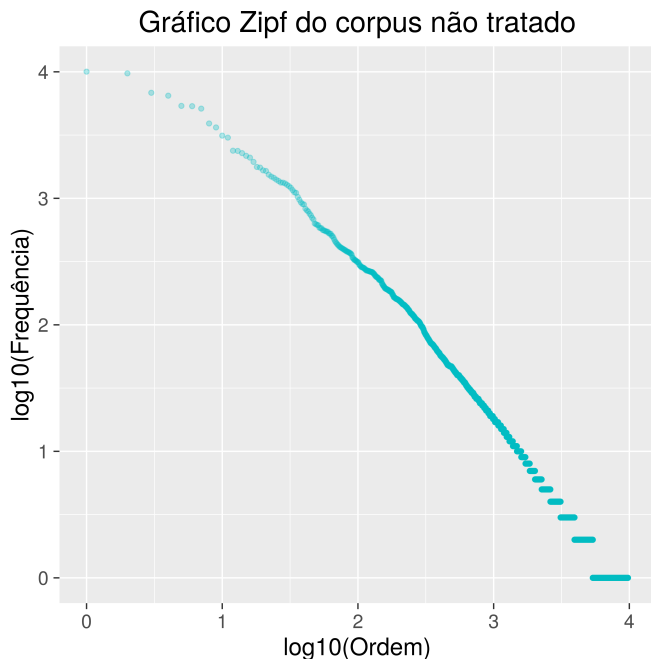
Como esperado, a distribuição das palavras do *corpus* segue a lei de Zipf, mas com um pequeno afastamento da linearidade para as palavras mais frequentes. Isso acontece porque as palavras mais frequentes são principalmente artigos, advérbios e conjunções com frequência de utilização parecida. A descontinuidade na parte superior do gráfico se deve à pequena compressão operada pela transformação logarítmica nos valores dos primeiros ranques. O mesmo acontece na parte inferior do gráfico, mas dessa vez a compressão reduzida se deve aos valores de frequência mais baixos. O gráfico permite verificar que as duas primeiras palavras, por exemplo, tem



**Figura 4.2:** Gráfico da função de probabilidade cumulativa empírica do logaritmo de base 10 da contagem de palavras por turno de fala do *corpus* não tratado. A linha vermelha indica o analisando e a linha azul indica o analista.

ambas contagens de cerca de  $10^4$ , ou seja 10.000 ocorrências cada uma em todo o *corpus*, o que equivale a quase 10% do total.

50% de todo o *corpus* corresponde às 47 palavras mais frequentes. Como pode ser visualizado na Tabela 4.2, a maioria absoluta das palavras é de pouco valor semântico: conjunções, artigos, preposições, advérbios, verbos auxiliares. O termo “ret3” é um marcador utilizado para indicar interrupções nas associações do analisando. O motivo para exposição dessas informações é deixar evidente que qualquer análise de linguagem natural irá enfrentar problemas com a escassez de dados. No nosso caso, a análise de *stopwords*, termos que cumprem principalmente uma função sintática mas não carregam muito valor semântico em si, irá remover a maior parte desses termos, reduzindo nosso *corpus* a cerca de um terço do tamanho



**Figura 4.3:** Gráfico de Zipf do *corpus* não tratado

original. Como a eliminação dos termos raros também irá reduzir o vocabulário mais ou menos na mesma proporção, as características distribucionais do *corpus* não são gravemente afetadas.

As informações descritivas do *corpus* foram processadas com auxílio do pacote `tm` (Feinerer & Hornik, 2015) e `ggplot2` (Wickham, 2009), ambos disponíveis para linguagem de programação R. Convém notar que a análise exploratória do *corpus* se foca, neste capítulo, quase que exclusivamente na dimensão quantitativa. Uma apreciação mais qualitativa dos dados do *corpus*, em conjunto com uma breve caracterização clínica do estudo de caso, é feita no próximo capítulo, em conjunto com a exploração dos resultados do modelo.

**Tabela 4.2:** Frequência absoluta e relativa das 47 primeiras palavras do *corpus* não tratado (com algumas palavras excluídas ao acaso por questão de espaço)

	Palavra	Ocorrências	Proporção
2	que	9.722	0,044
3	e	6.829	0,031
4	não	6.479	0,029
5	de	5.379	0,024
6	a	5.352	0,024
7	o	5.119	0,023
8	ret3	3.906	0,018
9	é	3.634	0,016
10	para	3.130	0,014
11	mas	3.021	0,014
12	com	2.380	0,011
13	um	2.370	0,011
14	porque	2.277	0,010
15	ela	2.168	0,010
16	ele	2.098	0,009
17	se	1.941	0,009
18	em	1.770	0,008
19	isso	1.750	0,008
20	uma	1.663	0,008
22	no	1.538	0,007
23	estava	1.486	0,007
25	como	1.406	0,006
26	mais	1.372	0,006
27	na	1.332	0,006
28	da	1.331	0,006
29	lá	1.313	0,006
30	coisa	1.280	0,006
31	por	1.249	0,006
32	era	1.212	0,005
33	bem	1.161	0,005
34	foi	1.116	0,005
35	sei	1.105	0,005
36	quando	1.027	0,005
42	só	803	0,004
44	aí	756	0,003
45	também	734	0,003
46	está	701	0,003
48	Total	111.040	0,502

## 4.2 Tratamento dos dados

### 4.2.1 Identificação de colocações

Para iniciar o tratamento do *corpus*, identificamos palavras que tendem a ocorrer uma na sequência da outra com probabilidade muito maior do que a esperada para os dados do *corpus*. Esses conjuntos de palavras que coocorrem com grande frequência e dentro de um contexto curto são chamadas de *colocações* e permitem identificar locuções importantes utilizadas pelos interlocutores. Utilizando como critério a estatística  $G^2$ , baseada na razão de verossimilhança entre as contagens esperadas a partir da distribuição marginal dos termos e a contagem observada no *corpus*, tal como implementada no pacote **quanteda** (Benoit & Nulty, 2016), identificamos as principais colocações baseadas em trigramas e bigramas, ou seja, três ou duas palavras em sequência, respectivamente. Apresentamos essas colocações em conjunto com sua contagem, excluindo aquelas compostas apenas por palavras de pouco conteúdo semântico e com menos de 50 ocorrências, na Tabela 4.3.

**Tabela 4.3:** Colocações baseadas em trigramas e bigramas

Trigramas	Contagem	Bigramas	Contagem
“eu não sei”	408	“eu não”	1.678
“eu acho que”	344	“não sei”	697
“é como se”	212	“eu acho”	477
“de alguma forma”	104	“como se”	420
“um monte de”	99	“sei lá”	354
“eu não lembro”	94	“alguma coisa”	240
“mais ou menos”	92	“às vezes”	216
“não me lembro”	91	“de repente”	191
“por um lado”	88	“pelo menos”	154
“como se eu”	83	“está bem”	111
“como se fosse”	77	“não lembro”	109
“por outro lado”	64	“pode ser”	108
“quando eu era”	64	“ano passado”	105
“fim de semana”	62	“outras coisas”	100
“eu não consigo”	57	“de novo”	86
“eu não tenho”	56	“por exemplo”	82
“eu não tinha”	56	“para que”	76
“eu estava lá”	55	“outras pessoas”	71
“não sei porque”	54	“nesse sentido”	64

É curioso notar que parte das locuções de trigramas encontradas pela busca das colocações se refere principalmente a negações, muitas delas associadas à incapacidade de analisando realizar uma ação. Algumas dessas ações, inclusive, parecem refletir o que lhe é demandado pela associação livre. A natureza das colocações deixa evidente o valor de utilizá-las na análise do *corpus*: muitas dessas palavras, ao serem isoladas, teriam pouco sentido ou seriam mesmo excluídas da versão final. Com sua identificação, poderemos investigar melhor os contextos de ocorrências dessas locuções para melhor interpretar seu papel nas associações livres.

As colocações baseadas nos bigramas, por outro lado, parecem estar relacionadas a locuções usuais do uso cotidiano da língua, com a possível exceção das hesitações indicadas por “não sei” e “eu não”. Apesar de haver inicialmente uma contagem alta de colocações baseadas bigramas, essa frequência cai rapidamente porque a maioria dos colocações encontradas envolvem palavras de pouco valor semântico que muitas vezes se seguem na construção de uma frase.

### 4.2.2 Lematização

Como definimos na formalização de nosso modelo de tópicos para as associações livres, a unidade básica que compõe as associações são os termos em sua forma canônica. Essa forma canônica é tradicionalmente denominada de *lema*, e o processo da retirada das inflexões de uma palavra é chamado de *lematização*. Um problema crucial do processo de lematização é a ambiguidade na forma escrita das palavras. Termos com significados diferentes mas escritos da mesma forma não são facilmente diferenciáveis, o que torna a utilização de um dicionário de lemas pouco útil para a tarefa.

Uma solução para contornar essa dificuldade é utilizar algoritmos que fazem uso de informações contextuais, como palavras vizinhas, para determinar melhor a qual lema corresponde a palavra num documento. Nós utilizamos um algoritmo desenvolvido para identificação da classe gramatical de palavras por meio de árvores de decisões probabilísticas denominado de *TreeTagger* (Schmid, 1995), reconhecido por sua alta acurácia. Seus dados de treinamento originais são de língua inglesa e alemã, mas há uma definição de treinamento realizada a partir de um grande *corpus* de língua portuguesa (Gamallo & Garcia, 2013). Como não há uma interface do *TreeTagger* dentro da linguagem de programação R, desenvolvemos uma rotina que aplica o *TreeTagger* aos documentos de uma coleção em língua portuguesa e retorna os lemas correspondentes mais prováveis. Especificamos, nessa rotina, para que o processo de lematização preservasse numerais e o gênero dos pronomes pessoais. No caso do *TreeTagger* não conseguir encontrar um lema para uma palavra (especialmente no caso das colocações apresentadas acima), ele deve retornar o termo original, sem modificações.

O processo de lematização não afeta o número de palavras por turno de fala nem o número de turnos de falas em cada sessão. O resultado esperado é uma diminuição do tamanho do vocabulário e uma consequente redução da quantidade de hapaxes, uma vez que diferentes palavras, ao perderem suas inflexões, são fundidas num mesmo lema. E, de fato, o vocabulário total foi reduzido de 9.768 para 5.225 termos, uma compressão de 46% – isso levando em consideração a criação de novos termos em função da identificação de colocações. O número de hapaxes foi reduzido de 4.391 para 1.890 termos, diminuindo a proporção de termos únicos no *corpus* para 36% do vocabulário.

Deixaremos para retomar as explorações feitas na Seção 4.1 após a realização do próximo passo, para podermos comparar mais



diretamente as descrições do *corpus* em seu estado bruto e após o tratamento para seu uso no modelo.

### 4.2.3 Remoção de *stopwords*

A última etapa de preparação do *corpus* é a identificação e remoção de *stopwords*, termos que possuem função eminentemente sintática e carregam pouco valor semântico por si só. Uma característica importante das *stopwords* é a sua generalidade: elas tendem a aparecer em qualquer contexto, e não possuem um perfil distribucional informativo. Por esse motivo, sua remoção não implica numa perda significativa, pois sua manutenção no *corpus* ou seria inócua, na medida em que esses termos não têm um padrão de coocorrência relevante para ser capturado pelo modelo; ou seria negativa, na medida em que sua alta frequência em vários contextos diferentes poderia distorcer a distribuição dos termos com maior peso semântico.

Os termos considerados *stopwords* são, em geral, definidos numa lista genérica para cada língua. O pacote `tm` possui uma lista de 203 termos considerados *stopwords* na língua portuguesa. Utilizamos esse material como referência, mas não o aplicamos diretamente. Nossa estratégia para identificação das principais *stopwords* se baseou em seu caráter distribucional. Primeiramente, normalizamos a contagem de cada termo em um documento (lembrando que, no nosso caso, cada turno de fala é um documento distinto) utilizando o total de termos presentes no documento. Com isso, cada termo é transformado em uma proporção. Essa normalização é necessária porque esperamos contagens mais altas de termos num documento mais longo, e, se utilizarmos apenas a frequência absoluta, os resultados serão influenciados por essa diferença. Termos pouco informativos terão como característica uma magnitude relativamente grande e com pouca variação de um documento para outro. Como contra-exemplo, imaginemos um termo que aparece com alta frequência num documento específico, mas com frequência relativamente menor em outros. Esse termo provavelmente será específico dos tópicos abordados no documento em que aparece com frequência, e portanto será relevante mantê-lo na análise.

Na prática, essa medida de relevância de um termo pode ser sumarizado pelo *coeficiente de variação* de cada termo, de acordo com a equação Equação 4.1. O coeficiente de variação é simplesmente o desvio-padrão de um termo dividido por sua média. Os termos com menor coeficiente de variação são justamente aqueles

menos informativos em termos distribucionais e, portanto, candidatos a serem removidos como *stopwords*.

$$cv_t = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}}{\frac{1}{N} \sum_{i=1}^N x_i} \quad (4.1)$$

Para não ficarmos limitados a apenas um critério, utilizamos também o valor de *idf* – *Inverse Document Frequency* – de cada termo. A medida *idf* é o logaritmo da razão entre o número total de documentos e o número de documentos nos quais o termo ocorre, conforme formalizado na Equação 4.2, e números mais baixos também indicam palavras que aparecem numa proporção grande de documentos e podem ser, portanto, pouco informativas.

$$\text{idf}_t = \log \left( \frac{|D|}{|\{d \in D : t \in d\}|} \right) \quad (4.2)$$

Os resultados entre as duas estratégias foi bastante parecido: os termos com pontuação mais baixa são praticamente os mesmos, com alguma pequena mudança na ordem. Para delimitar quais palavras seriam excluídas, selecionamos as duzentas primeiras obtidas pelos dois procedimentos, removemos os termos duplicados e avaliámos manualmente a pertinência de cada um dos termos restantes. Uma lista final de 70 termos foi definida como o conjunto de *stopwords* do *corpus* e cada ocorrência desses termos foi removido. Foram também removidos os hapaxes e termos que apareceram num único documento, uma vez que, apesar de constituídos de palavras de forte conteúdo semântico, suas propriedades distribucionais são fracamente informativas para estabelecer relações de coocorrência claras com os outros termos do *corpus*.

#### 4.2.4 Descrição do *corpus* tratado

Em função da remoção das palavras com pouco conteúdo semântico, 47 turnos de fala foram removidos da versão final do *corpus*. Além de comporem uma proporção pequena, os turnos de fala desconsiderados não continham nenhuma palavra significativa, o que torna sua remoção pouco impactante no ajuste do modelo. Restaram 3.172 turnos de falas, dos quais 1.599 são do analista e 1.573 são do analisando. Apesar da redução, o número médio de turnos de fala por sessão continuou praticamente o mesmo, 67.

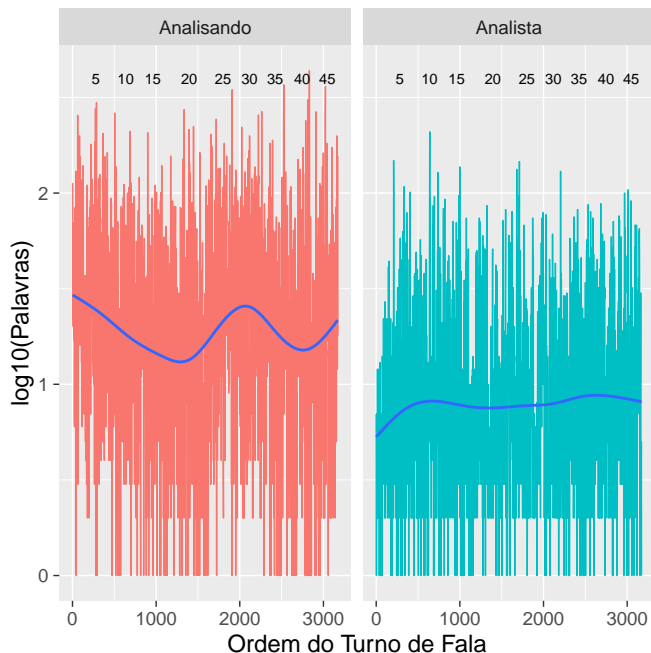
O vocabulário final sofreu uma redução drástica. Com a eliminação dos hapaxes e das *stopwords*, o número total de termos foi reduzido a menos de um terço do original, 3.007. A restrição no vocabulário também impactou o número de ocorrências dos termos no *corpus*: das 221.187 palavras do *corpus* não tratado, foram retidas 82.457, 37% do total original. Esse aparente desperdício de dados é justificado para a obtenção de um modelo que relacione termos com relevância semântica; mesmo assim, reflete um grande problema no tratamento quantitativo de dados textuais: a maior parte dos dados tem pouco valor para a análise, implicando numa quantidade esparsa de informações que são custosas de serem obtidas.

**Tabela 4.4:** Informações descritivas do *corpus* tratado.

	Analizando	Analista	Total
Turnos de Fala	1.573	1.599	3.172
Média de TdF por Sessão	33	34	67
Mediana de TdF por Sessão	33	34	67
Vocabulário	2.573	2.321	3.007
Palavras	58.648	23.809	82.457
Média de Palavras por TdF	37	15	26
Mediana de Palavras por TdF	21	7	12

O resultado é que os documentos analisados foram consideravelmente reduzidos. A média de palavras nos turnos de fala do analisando ficou em 38 (DP=48), contra 15 (DP=20) palavras nos turnos de fala do analista. A avaliação da contagem de palavras em cada turno de fala, porém, não sofreu nenhuma modificação significativa. Ainda há uma grande variação no número de palavras empregado em cada turno de fala tanto para analisando quanto para analista, e a tendência temporal desses valores segue o mesmo padrão de estabilidade, no caso do analista, e a mesma flutuação ao longo do tempo, para o analisando, conforme a Figura 4.4.

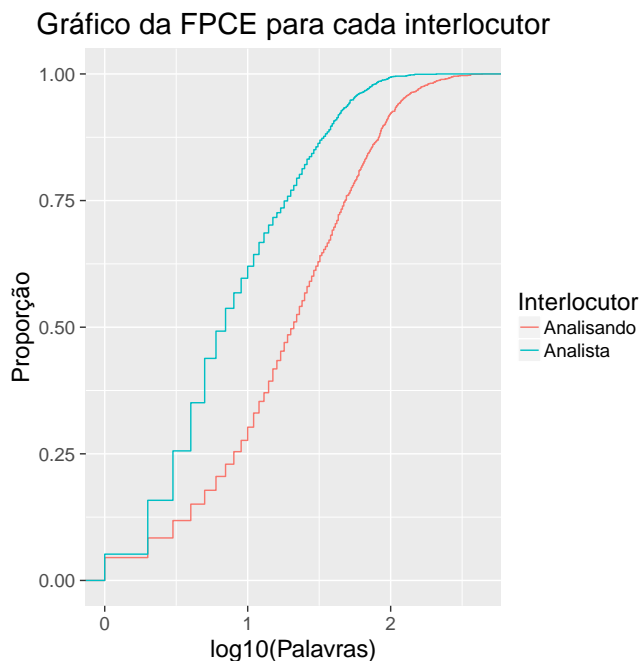
O gráfico da função de probabilidade cumulativa empírica para o *corpus* tratado é apresentado na Figura 4.5. Como já visto nas estatísticas sumárias, houve uma considerável redução no número de palavras por turno de fala, tanto para analisando quanto analista. Para o analisando, a proporção de documentos com menos de 10 palavras é de 25%, enquanto que para o analista essa proporção chega a 62%. Os turnos de fala dentro do intervalo interquartilício do analisando possuem de 10 a 50 palavras; no caso do



**Figura 4.4:** Gráfico de linha do logaritmo da contagem de palavras em cada turno de fala para o *corpus* processado. Sobre cada gráfico foi sobreposto uma linha de regressão não-linear para facilitar a visualização das tendências temporais.

analista, esse intervalo vai de 3 a 18.

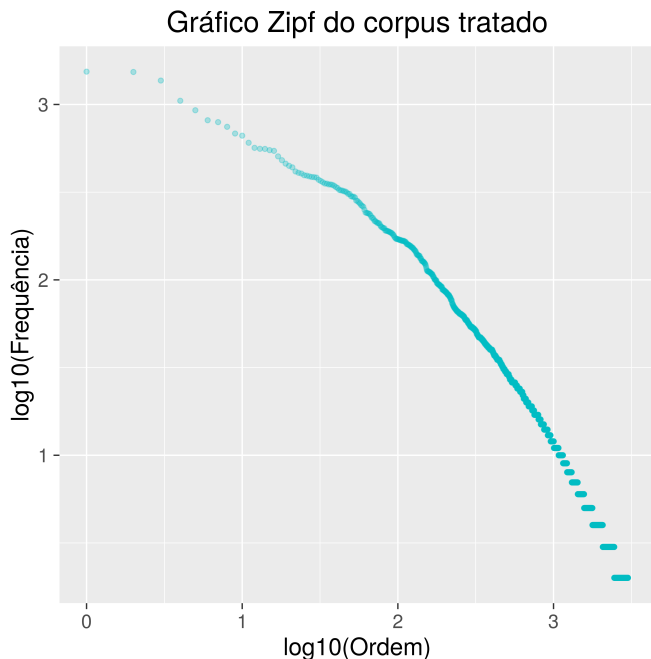
O gráfico Zipf do *corpus* tratado apresenta um claro desvio da linearidade, em contraste com o comportamento esperado e o verificado para o *corpus* bruto. Uma análise mais detalhada da Figura 4.6 permite verificar a existência de duas tendências lineares na frequência dos termos. Os termos menos frequentes, aqueles com frequência até cerca de  $10^{2.25}$  ( $\sim 178$ ), possuem uma inclinação negativa mais forte, sugerindo uma diferença maior entre frequência de termos com ordens próximas. Os termos com frequência acima de  $10^{2.25}$  possuem uma inclinação mais suave, indicando uma diferença menor entre termos com ordens próximas – lembrando, é claro, que o gráfico apresenta a frequência absoluta em escala logarítmica. A presença de duas inclinações diferentes é um resultado esperado do



**Figura 4.5:** Gráfico da função de probabilidade cumulativa empírica do logaritmo de base 10 da contagem de palavras por turno de fala no *corpus* tratado. A linha vermelha indica o analisando e a linha azul indica o analista.

tratamento do *corpus*, uma vez que os termos que mais sofrem impacto da lematização e da remoção de *stopwords* são aqueles com maior frequência. Portanto, a verificação gráfica da frequência das palavras indica uma suavização na diferença entre os termos mais frequentes, um resultado desejável por impedir que palavras excessivamente frequentes distorçam a distribuição de tópicos.

Uma das consequências dessa suavização é que 50% do *corpus* tratado correspondem aos 124 termos mais frequentes. Uma diferença considerável em comparação com o *corpus* bruto, no qual essa proporção correspondia às 47 primeiras palavras. A título de completude, a Tabela 4.5 apresenta os 20 termos mais frequentes da versão final do *corpus*.



**Figura 4.6:** Gráfico de Zipf do *corpus* tratado.

### 4.3 Seleção do modelo

Apesar de já termos o modelo formalmente definido, inclusive em termos de sua implementação prática, ainda é preciso escolher um ajuste adequado aos dados do *corpus* tratado. O ajuste de um mesmo modelo de tópicos a um conjunto de dados é influenciado por dois elementos principais. O mais importante e com impactos práticos é o número de tópicos, que precisa ser definido *a priori* – a não ser no caso dos modelos que tentam estimar esse número por meio de processos de Dirichlet (Blei et al., 2010), mas não é o caso de nosso modelo nem do algoritmo do STM. O segundo aspecto que influencia o ajuste do modelo, de caráter computacional, é o problema da multimodalidade (Roberts, Stewart, & Tingley, 2016). Os modelos de tópicos, de maneira geral, possuem várias soluções mais ou menos parecidas (*modas*) e, como os algoritmos para encontrar essas soluções são aproximados em função da intratabilidade

**Tabela 4.5:** Frequência absoluta e relativa das 20 primeiras palavras do *corpus* tratado.

	Palavra	Ocorrências	Proporção
1	falar	1.539	0,019
2	ficar	1.532	0,019
3	dizer	1.369	0,017
4	bem	1.050	0,013
5	eunão	927	0,011
6	ver	813	0,010
7	pensar	793	0,010
8	poder	746	0,009
9	querer	683	0,008
10	bom	664	0,008
11	lembrar	605	0,007
12	parecer	566	0,007
13	aqui	559	0,007
14	agora	558	0,007
15	gente	549	0,007
16	depois	544	0,007
17	talvez	506	0,006
18	sentir	481	0,006
19	pouco	460	0,006
20	tudo	447	0,005

da distribuição posterior, inicializações diferentes podem produzir soluções diferentes para um mesmo modelo e conjunto de dados, às vezes com poucas alterações, às vezes com alterações substanciais.

#### 4.3.1 Número de tópicos

A não ser em raros casos em que o número de tópicos pode ser escolhido a partir de estudos prévios ou de conhecimentos sobre o *corpus* em estudo, o problema de sua definição só pode ser resolvido de maneira aproximativa. Não é possível ignorar as dificuldades e simplesmente escolher um número ao acaso ou pouco informado, pois o impacto será significativo no resultado do ajuste do modelo final. Quando o número de tópicos é pequeno demais, a estrutura descoberta tende a ser excessivamente genérica e pouco informativa, reunindo termos pouco relacionados num mesmo tópico. Se, pelo contrário, o número de tópico é excessivo, a granularidade da estru-

tura descoberta se torna muito específica e também pouco informativa, pois as distribuições irão refletir informações excessivamente contextuais. Mesmo assim, não existe um único número exato de tópicos para cada *corpus*, e um intervalo considerável pode trazer resultados próximos e satisfatórios.

Utilizamos três estratégias para selecionar o valor mais adequado. A primeira delas, mais comum na literatura de modelos de tópicos, avalia a *verossimilhança* de um conjunto de dados retidos aleatoriamente e não utilizados na estimação para cada modelo ajustado (Griffiths & Steyvers, 2004). Essa estratégia, baseada na validação cruzada, é comum na área de aprendizagem de máquina em função de sua fácil interpretação. O modelo que atribuir maior verossimilhança aos dados retidos é o que melhor generaliza para dados não observados e, portanto, seu número de tópicos deve ser mais adequado, pelo menos com relação aos modelos comparados. Apesar de sua interpretação estatística ser justificada, a verossimilhança de dados retidos pode não representar muito bem a utilização do modelo de tópicos por seres humanos, porque não leva em consideração nenhuma informação sobre a interpretabilidade dos tópicos encontrados. Além disso, a validação cruzada é computacionalmente intensiva e suscetível às variações nos dados deixados à parte do processo de estimação (Roberts, Stewart, & Airolidi, 2016).

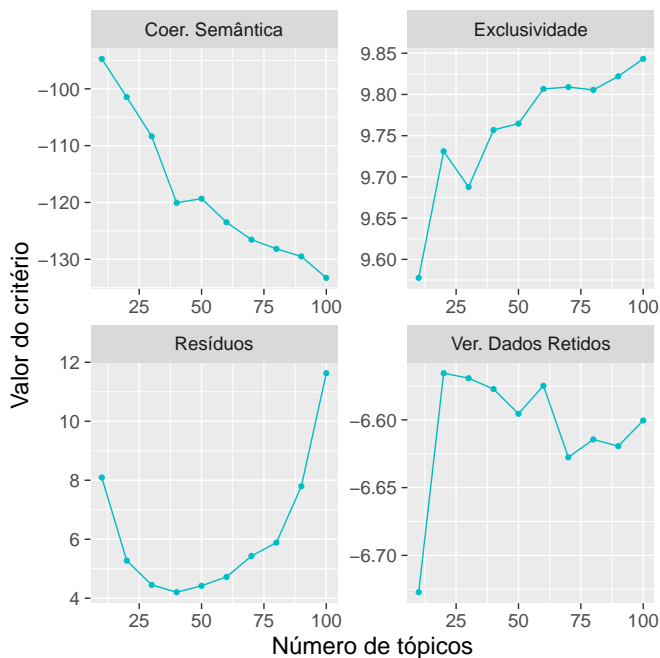
A validação cruzada por dados retidos é implementada de uma maneira peculiar no pacote **stm**. Em vez de reter documentos inteiros, a função **searchK** retém 50% das palavras em um subconjunto dos documentos do *corpus* não maior do que 10% do número total. Após ajustar modelos com diferente número de tópicos, a função calcula a verossimilhança das palavras deixadas de fora com base nos parâmetros de proporção de tópicos de cada documento (Roberts, Stewart, & Airolidi, 2016).

Roberts, Stewart, e Tingley (n.d.) também implementam uma segunda estratégia no pacote **stm**, baseada não em critérios estatísticos abstratos como a verossimilhança de dados deixados à parte, mas em duas métricas ligadas à interpretabilidade dos tópicos descobertos por seres humanos. A *coesão semântica* é uma medida da efetiva coocorrência, dentro dos documentos, das palavras com alta probabilidade dentro de um mesmo tópico. A *exclusividade*, por sua vez, indica se as palavras mais importantes de um tópico têm baixa probabilidade de aparecer como importantes em outro tópico (Roberts et al., 2014). Um ajuste ideal possui vários tópicos com coesão semântica e exclusividade altas.



A terceira estratégia é baseada nos pressupostos do modelo. Taddy (2011) demonstra que, sob o modelo correto, a variância dos resíduos deve ser próximo a um. Valores acima de um podem indicar que o número de tópicos escolhido é insuficiente. Na prática, porém, como o modelo ajustado nunca é propriamente correto, uma vez que se trata de uma aproximação idealizada do processo gerador de dados, dificilmente obtemos uma variância muito próxima a um. Mesmo assim, o critério permanece útil ao compararmos vários modelos entre si: aquele com menor variância residual estará mais próximo das definições pressupostas pelo modelo.

Utilizando a função `searchK` do pacote `stm`, ajustamos dez modelos com número de tópicos variando de 10 a 100, espaçados de 10 em 10. A definição de cada modelo foi exatamente igual, salvo o número de tópicos. Os resultados podem ser conferidos na Figura 4.7.



**Figura 4.7:** Verossimilhança de dados retidos, resíduos, exclusividade e coerência semântica do modelo ajustado em função do número de tópicos.

Os resultados não são particularmente favoráveis a um único número de tópicos. Pelo critério da coerência semântica, modelos com menos tópicos apresentaram um melhor desempenho – o que é esperado, por um lado, porque o número reduzido de tópicos força a coexistência de termos distintos dentro de um mesmo tópico e, portanto, há uma chance maior de eles coocorrerem nos documentos. O resultado do critério da exclusividade aponta para a direção contrária: modelos com mais tópicos têm um escore mais alto, justamente porque eles separam melhor os termos em tópicos diferentes, diminuindo a probabilidade de um mesmo termo aparecer com alta probabilidade em dois tópicos. Utilizando os dois critérios em conjunto, um número entre 40 e 60 parece equilibrar coerência com exclusividade.

Como esperado, nenhum dos modelos apresentou variância residual próxima de um. A menor variância pertence ao modelo com 40 tópicos, seguido de perto pelo modelo com 50. Por fim, a verossimilhança dos dados retidos varia pouco nos modelos de 20 a 60 tópicos, com uma pequena queda para o modelo com 50. Em seguida, a verossimilhança dos dados retidos diminui drasticamente, melhorando pouco até o modelo com 100 tópicos. Novamente, a faixa de 40 a 60 tópicos parece ser a mais indicada. Esse número, indicado empiricamente a partir dos critérios elencados, está de acordo com o que é geralmente recomendado na literatura sobre modelos de tópicos para *corpora* com mais de mil mas menos de dez mil documentos (Blei & Lafferty, 2009; Roberts et al., 2013).

### 4.3.2 O problema da multimodalidade

A distribuição posterior de qualquer modelo de tópicos é multimodal, como qualquer modelo com variáveis latentes. Isso significa que, para um mesmo modelo com especificações fixas e um mesmo conjunto de dados, os algoritmos que estimam os parâmetros por meio de critérios de otimização podem encontrar resultados diferentes se iniciarem a exploração do espaço da distribuição posterior em locais diferentes. A multimodalidade torna quase impossível encontrar uma solução global ótima, e precisamos nos contentar com máximos locais que possuam um bom desempenho. Há duas principais fontes de multimodalidade: a troca do número que identifica os tópicos, uma vez que não há nenhum significado intrínseco associando tópico e seu número no modelo; e a presença de variáveis latentes tanto para identificar a proporção de tópicos quanto para

a probabilidade de termos dentro dos tópicos (Roberts, Stewart, & Tingley, 2016).

Não há nenhuma solução simples para contornar o problema da multimodalidade e encontrar a moda máxima global. Uma solução por força bruta é estimar o mesmo modelo diversas vezes com diferentes valores de inicialização. Ao fim, os modelos ajustados são comparados e é feita a escolha do modelo com maior verossimilhança – ou qualquer outro critério desejável, como a interpretabilidade dos tópicos obtidos. Se o *corpus* é consideravelmente informativo, ou seja, se ele é composto por um número grande de documentos e palavras, e as palavras aparecem de maneira razoavelmente bem separada, as soluções das diferentes modas tendem a ser bastante semelhantes, pelo menos para os tópicos mais facilmente identificáveis.

Uma alternativa recentemente proposta por Roberts, Stewart, e Tingley (2016) conjuga técnicas de álgebra linear com estimação variacional para obter uma solução determinística e com maior probabilidade de se encontrar próximo da moda máxima global. Em vez de inicializar o algoritmo de ajuste do modelo aleatoriamente ou por meio de simulações preliminares, Roberts, Stewart, e Tingley (2016) operam a *decomposição espectral* da matriz de termos e documentos. Essa decomposição é uma fatoração não-negativa da matriz original que permite recuperar os valores esperados de variáveis latentes diretamente a partir da matriz de contagem. A estratégia da decomposição espectral é pressupor que há no *corpus* termos bem separados, ou seja, que pertencem exclusivamente a um único tópico. Após identificar esses termos, o pertencimento dos outros termos aos tópicos latentes pode ser deterministicamente computado. Na implementação do algoritmo do STM, a inicialização por decomposição espectral relaxa o pressuposto de termos pertencente a um único tópico, pois o resultado da decomposição é utilizado apenas para dar as coordenadas iniciais que serão otimizadas por meio de maximização expectância variacional.

Nos experimentos apresentados por Roberts, Stewart, e Tingley (2016), a inicialização por decomposição espectral, além de resultar sempre no mesmo modelo ajustado, apresenta também verossimilhança marginal consideravelmente maior do que qualquer modelo obtido por meio de outras inicializações. Por esse motivo, escolhemos ajustar nosso modelo final fazendo uso da inicialização espectral.

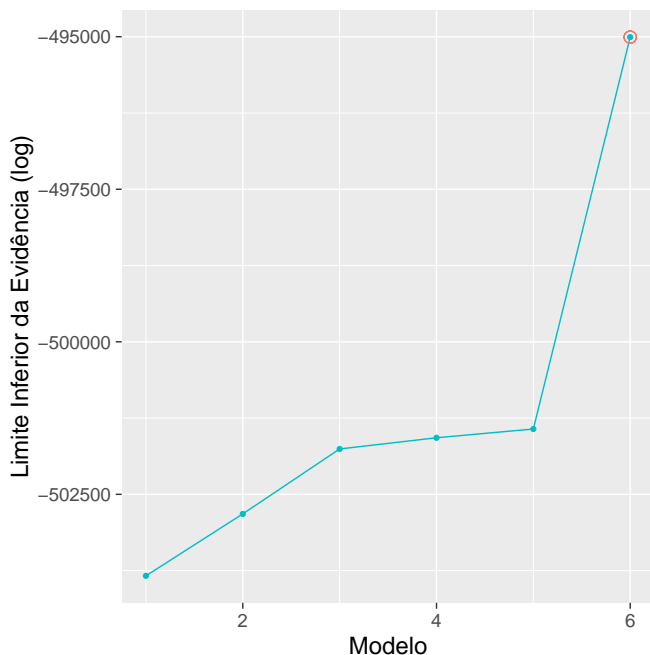
### 4.3.3 Decisão final

Tendo em vista os resultados e considerações feitas nesta seção, optamos por ajustar um modelo com 50 tópicos, valor intermediário dentro da faixa identificada como mais relevante. Optamos por utilizar a estratégia de inicialização espectral como forma de evitar a necessidade de ajustar vários modelos diferentes.

Para certificar que o ajuste encontrado pela inicialização espectral não é necessariamente pior do que o ajuste por outras estratégias, suscetíveis à variação de acordo com a inicialização, comparamos o limite inferior da evidência (*evidence lower bound*) obtido pelo modelo escolhido com outros cinco ajustados por meio de inicialização por simulações do algoritmo de Gibbs para a LDA. O limite inferior da evidência é uma estimativa aproximada do denominador da regra de Bayes, ou seja, a probabilidade dos dados observados sob o modelo em consideração, e pode ser utilizado para seleção de modelos. Como pode ser visto na Figura 4.8, o modelo escolhido possui um limite inferior da evidência consideravelmente maior do que os outros modelos obtidos por inicialização por simulações.

Para recapitular, aproximamos o modelo de tópicos para as associações livres, proposto no Capítulo 3, utilizando a implementação do *Structural Topic Model* do pacote `stm`, para linguagem de programação R, utilizando os seguintes parâmetros:

- O número de tópicos,  $K$ , definido como 50, em função do processo de seleção apresentado acima;
- O parâmetro de concentração  $\kappa$ , que governa a distribuição de termos nos complexos, definido para o padrão de 0.1, valor utilizado para a obtenção de tópicos esparsos (com poucos termos com alta probabilidade e vários termos com baixa probabilidade);
- Os parâmetros  $a$  e  $b$ , que governam o grau de regularização sobre os coeficientes do modelo de regressão multinomial, foram definidos para o valor padrão de 1, de forma obter estimativas fracamente regularizadas;
- O modelo de regressão multinomial, que define a proporção de tópicos sobre documentos, ficou definida como  $\vec{\mu}_d = \vec{\gamma}^{inter} x_d^{inter} + \vec{\gamma}^{sess\tilde{a}o} S(x_d^{inter})$ , ou seja, com um coeficiente para o interlocutor (por tópico) e um conjunto de coeficientes (também por tópico) para a *spline* cúbica aplicada sobre o número da sessão;



**Figura 4.8:** Limite inferior da evidência para seis modelos de 50 tópicos. O modelo escolhido está marcado em vermelho.

- Por fim, a estratégia de ajuste utilizou a inicialização espectral, pelos motivos apontados acima.

## 4.4 Análise do ajuste aos pressupostos

Os modelos de tópicos, por serem de invenção relativamente recente, não possuem ainda procedimentos canônicos para avaliação de seu ajuste, como é o caso, por exemplo, da análise de resíduos nos modelos lineares. Seguiremos aqui a sugestão de Mimno e Blei (2011) de utilizar a *checagem preditiva posterior* (*posterior predictive check*), uma estratégia desenvolvida no contexto da estatística bayesiana para detectar discrepâncias relevantes entre os resultados obtidos pelo modelo, os dados observados e os pressupostos (Gelman et al., 2014). Na checagem preditiva posterior, replicações do conjunto de dados são obtidas por meio de simulações da distribuição poste-

rior. O resultado é uma distribuição preditiva de novos resultados com base nas informações aprendidas pelo modelo. Após gerada, a distribuição preditiva pode, então, ser contrastada com aspectos importantes dos dados originais ou dos pressupostos do modelo. A checagem preditiva posterior pode ser considerada uma forma de validação interna do modelo, ou seja, uma avaliação da adequação das predições do modelo às características consideradas relevantes nos dados utilizados no ajuste.

Para elaborar uma checagem preditiva posterior, é necessário estabelecer uma *estatística de teste*, uma função que resuma características julgadas importantes sobre os dados. Há várias estatísticas relevantes que podem ser computadas a partir de um *corpus* textual, mas nos focaremos numa relação fundamental pressuposta pelo modelo. Tanto na LDA quanto no STM é feita a suposição de que, uma vez definido um tópico  $z$  dentro de um documento, a escolha das palavras se dá por sorteio aleatório da distribuição de tópicos,  $\beta_z$ . Como já discutimos, essa pressuposição simplista contraria a relevância que a ordem das palavras apresentam na constituição do significado, mas é uma aproximação boa o suficiente e que facilita consideravelmente a computação do modelo – como no nosso caso, que tivemos de abdicar em alguns aspectos de nossa proposta original. A proposta de Mimno e Blei (2011) consiste em fazer uso dessa pressuposição para identificar o quanto o modelo, ao pressupor independência, se afasta dos dados reais. A intuição por trás dessa ideia é simples: além de avaliar quanto o modelo é inadequado, permite também identificar tópicos nos quais o pressuposto é bem atendido, o que os torna mais confiáveis.

A estatística de teste proposta é a informação mútua entre as variáveis  $D$  e  $W$ , que indicam os documentos e as palavras, respectivamente.

$$\text{MI}(W, D \mid z = k) = \sum_w \sum_d p(w, d \mid k) \log \frac{p(w, d \mid k)}{p(w \mid k)p(d \mid k)} \quad (4.3)$$

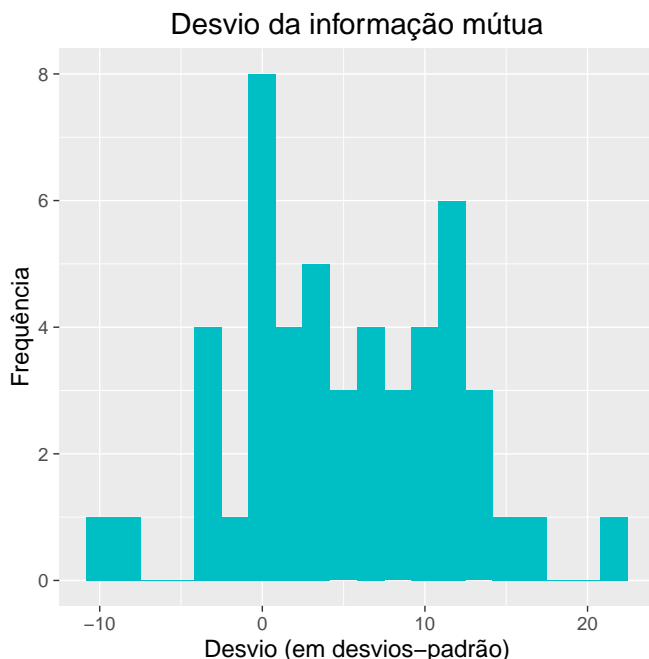
Condicionada à atribuição de uma palavra  $w_{d,n}$  a um tópico específico  $z_{d,n}$ , a variável  $W$  deve ser independente de  $D$ , pelas pressuposições do modelo. Portanto, a informação mútua entre  $D$  e  $W$  deve ser próxima de zero no limite de uma amostra infinita. Porém, como o modelo é ajustado a uma quantidade finita de dados, a informação mútua pode se distanciar de zero. A checagem preditiva posterior do modelo de tópicos consiste em verificar a informação

mútua predita pelo modelo em função da amostra finita e compara com os valores observados para os dados.

O pacote `stm` não fornece diretamente uma estimativa de  $z$  para cada palavra, mas esse valor pode ser facilmente obtido por meio de estimação MAP (*maximum a posteriori*) utilizando os outros parâmetros estimados pelo modelo:  $p(z_i | w_i, \theta_d, \beta) \propto p(z_i)p(w_i | z_i) = \theta_d \beta$ . Em 11 de 50 tópicos do modelo que selecionamos, a informação mútua observada a partir dos dados está dentro do intervalo de probabilidade de 95% predito pelo modelo. O desempenho pode parecer fraco, mas os estudos empreendidos por Mimno e Blei (2011) indicam que é usual as predições do modelo subestimarem a informação mútua presente nos dados. Em função dessa discrepância usual, eles sugerem utilizar uma medida de desvio para avaliar a qualidade dos tópicos entre si. Essa medida consiste na distância, em desvios-padrão, do valor observado em relação à média dos valores previstos. Tópicos com um escore de desvio muito alto destoam consideravelmente dos pressupostos do modelo, e aqueles com um desvio próximo a zero são mais confiáveis. A distribuição dos escores de desvio para cada tópico do modelo é apresentada na Figura 4.9.

Como esperado, o modelo subestima a informação mútua entre palavras e documentos. O desvio médio é de 5,2 desvios-padrão (DP = 6,7). Mesmo assim, metade dos tópicos está a cinco desvios-padrão da média predita pelo modelo, um valor consideravelmente baixo se nos basearmos nos valores obtidos por Mimno e Blei (2011) em seus experimentos, como 37,9 para um *corpus* de notícias e 44,4 para um *corpus* de blogs. Os resultados da checagem preditiva indicam que, apesar do modelo não capturar algumas dependências importantes entre palavras e documentos, seu desajuste é relativamente pequeno quando comparamos aos dados da literatura. De qualquer forma, para fundamentar as inferências nos resultados que estão mais próximos dos pressupostos do modelo, expomos no Capítulo 5 um subconjunto de tópicos que, além de serem semanticamente coerentes, apresentam uma desviância dentro dos limites aceitáveis, ou seja, com valor absoluto inferior a 3.

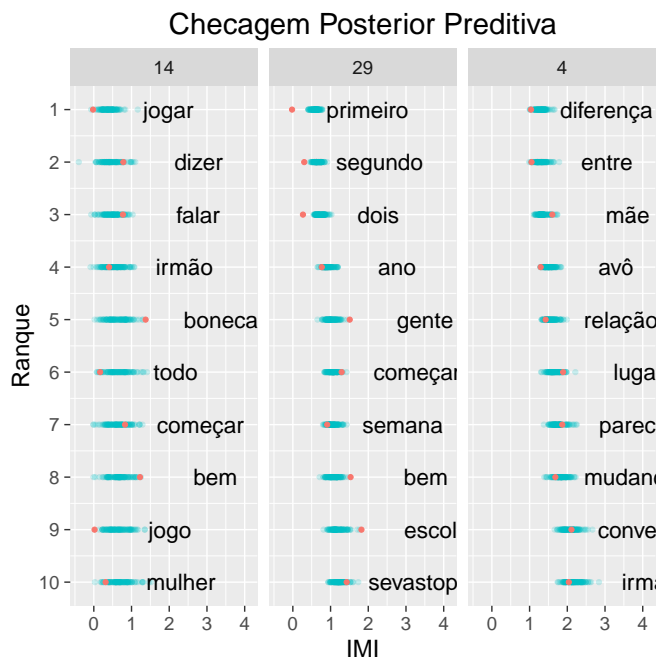
Mimno e Blei (2011) também sugere um teste baseado na informação mútua instantânea para avaliar desvios no âmbito das palavras individuais, mas não aprofundaremos nessa extensão da checagem aqui por julgar ser suficiente avaliar a qualidade dos tópicos de maneira geral. Mesmo assim, apresentamos os resultados da checagem preditiva para as dez primeiras palavras de três tópicos



**Figura 4.9:** Distribuição dos escores de desvio dos tópicos.

na Figura 4.10. Os três tópicos apresentados foram escolhidos por apresentar, respectivamente, a menor, a mediana e a maior informação mútua entre os tópicos. Os pontos pretos representam valores computados a partir de replicações baseadas no modelo, e os pontos vermelhos indicam a informação mútua instantânea realizada com os dados do *corpus*. Pelos pressupostos do modelo, a informação mútua calculada condicionalmente à atribuição de tópico deveria ser igual a zero. Num *corpus* finito, como já explicado, é esperado que esse valor se afaste um pouco de zero. Um tópico ideal, ajustado a um *corpus* grande o suficiente, deveria ter os valores simulados e os valores realizados sobrepostos e igual a zero para todos os termos. Numa situação real, é esperado que os valores estejam acima de zero, e os pontos simulados a partir do modelo dão uma ideia do intervalo aceitável. A principal avaliação a ser feita, portanto, é verificar a coincidência dos pontos simulados, em preto, e os pontos realizados, em vermelho.





**Figura 4.10:** Checagem Preditiva Posterior da Informação Mútua para os dez primeiros termos de três tópicos. Os tópicos escolhidos representam, da esquerda para a direita, tópicos com baixa, média e alta informação mútua entre palavras e documentos. Os pontos azuis representam a distribuição esperada da MI para cada palavra sob o modelo de tópicos. Os pontos vermelhos indicam a informação mútua calculada para os dados do *corpus*. Um tópico ideal teria os valores simulados e reais coincidindo e próximos a zero. O mais relevante, porém, é que os pontos simulados coincidem com os pontos efetivos.

O tópico 14, de menor informação mútua, apresenta uma variação considerável nos valores simulados, o que sugere que o ajuste poderia se beneficiar de um *corpus* maior. Mesmo assim, os valores realizados estão todos dentro do intervalo esperado, ainda que em alguns casos esteja no limite aceitável. O tópico 29, com informação mútua mediana, superestima a informação mútua instantânea para as palavras mais prováveis, mas também não destoa muito com relação aos valores realizados. O tópico 4, por fim, com maior informação mútua, também consegue capturar com confiança os valores realizados em seu intervalo esperado. A checagem preditiva posterior dos termos dentro dos tópicos sugere que o modelo ajustado está consideravelmente adequado à distribuição das palavras no *corpus*.

## 4.5 Validação do modelo por predição de dados de teste

Os procedimentos relatados até agora dão conta do tratamento pelo qual passou o *corpus*, os procedimentos iniciais para decidir a escolha do ajuste mais adequado do modelo aos dados e a consistência das estimativas com os pressupostos mais fortes. Antes de analisarmos os resultados obtidos pelo modelo, porém, é crucial fazê-lo passar por testes que nos permitam ter confiança nas conclusões. Blei (2012), criador da LDA, enfatiza a necessidade da utilização de critérios de validação para o uso de modelos de tópicos em áreas específicas, uma vez que sua natureza não-supervisionada não permite fazer inferências seguras para variáveis externas sem algum indício de validade. Esses critérios não podem ser definidos de uma vez por todas para qualquer aplicação dos modelos de tópicos e exigem conhecimento da área específica. Em nossa pesquisa, utilizamos dois procedimentos para obter indícios de validade do ajuste do modelo para o caso estudado: desempenho dos parâmetros do modelo como preditores dos metadados e predição das associações em testes de associação de palavras.

No contexto do uso de modelos de tópicos em estudos sobre psicoterapia, o principal critério de avaliação do desempenho do modelo é a predição de variáveis relacionadas às sessões clínicas, como identificação do interlocutor, tipo de psicoterapia e rótulos advindos da aplicação de escalas por juízes (Atkins et al., 2012, 2014; Imel, Steyvers, & Atkins, 2015b). Para a aplicação do modelo proposto

nesta tese, trabalhamos com duas variáveis que funcionam como metadados do *corpus*: o interlocutor a quem pertence o turno de fala e o número da sessão na qual ele ocorreu. Uma primeira proposta de validação do modelo é verificar se os parâmetros ajustados do modelo são informativos o suficiente a ponto de permitirem recuperar, sem nenhum outro dado adicional, informações externas aos dados textuais, como os metadados.

Formulando o problema de maneira mais clara: desejamos verificar se os parâmetros  $\theta_d$ , que o modelo de tópicos estima a partir dos dados textuais para cada documento e deveria, teoricamente, dar uma ideia do conteúdo de cada turno de fala, é suficiente para predizer o interlocutor e a sessão do tratamento com precisão, conforme esperado pelo modelo formal. Como a relação entre a proporção de tópicos dada por  $\theta_d$  e os desfechos a serem preditos não é linear, optamos pela utilização de um algoritmo capaz de capturar relações não-lineares com facilidade: o *random forests* (tal como implementado por Liaw & Wiener, 2002), baseado na construção de diversas árvores de regressão utilizando subconjunto dos preditores para melhorar a predição para dados fora da amostra de treinamento (James, Witten, Hastie, & Tibshirani, 2014). A utilização de *random forests* como estratégia de validação de modelos de tópicos já foi proposta no estudo Imel et al. (2015b), pela mesma justificativa proposta aqui.

Utilizamos a estratégia de *k-fold cross-validation* (validação cruzada pela divisão em  $k$  partes) para calcular o erro preditivo médio. Utilizamos o padrão de  $k = 10$ , o que significa que treinamos dois modelos de *random forests* – um para predição do interlocutor e outro para predição do número da sessão, e ambos tendo o vetor  $\theta_d$  como preditor para cada documento – utilizando dados de nove das dez partições aleatórias dos 3.172 documentos e testamos seu desempenho preditivo na partição deixada de fora do treinamento. Esse procedimento é reiterado para cada uma das dez partições. No caso da tarefa de identificação do interlocutor, o desempenho preditivo é mensurado pela *taxa de erro*, ou seja, pela soma de todos os documentos classificados erroneamente. Na tarefa de predição do número da sessão, por sua vez, o desempenho é medido pelo *erro quadrático médio* e pelo *erro absoluto mediano*. Por fim, calculamos a média dos valores do desempenho preditivo obtidos em cada partição de dados deixados de fora do treinamento (James et al., 2014).

A intuição desse procedimento é que, se os parâmetros do mo-

delo de tópicos, ajustados principalmente a partir dos dados textuais, permitem identificar o interlocutor e o momento do tratamento com precisão, isso significa que os resultados do ajuste estão coerente com o modelo formal e com critérios externos, sendo, portanto, confiáveis. Caso o desempenho preditivo das informações obtidas pelo modelo seja fraco, os resultados obtidos pelo modelo de tópicos são incoerentes e, assim, pouco confiáveis para quaisquer inferências.

No caso da predição do interlocutor com base na proporções de tópicos estimado pelo modelo, a taxa de erro médio por validação cruzada foi, surpreendentemente, zero. Isso significa que o modelo consegue separar com clareza o conteúdo utilizado nas intervenções do analista das associações livres enunciadas pelo analisando e, portanto, que os diferentes complexos identificados são informativos com relação aos dois interlocutores.

Para a predição do número da sessão à qual pertence cada turno de fala, o erro quadrático médio (*mean squared error*) por validação cruzada foi de 10,2. Aplicando a raiz quadrada sobre o valor para obtermos o valor na escala original, verificamos que as informações sobre a proporções de tópicos em cada turno de fala tende a errar, em média, por cerca de 3 sessões. Utilizando uma medida de desempenho mais robusta, o erro mediano absoluto, que é menos influenciado por valores excessivamente destoantes, obtemos um desempenho ainda melhor: em pelo menos 50% dos casos, o erro de predição foi de apenas uma sessão. Ambas as medidas de desempenho indicam que o parâmetro  $\theta_d$  é consideravelmente informativo sobre o momento em que se encontra o tratamento, sugerindo que o modelo de tópicos é capaz de capturar variações nos temas abordados ao longo do tempo, conforme esperado por sua definição formal.

## 4.6 Validação do modelo por teste de associação de palavras

Além do desempenho das estimativas do modelo na predição de critérios externos, fizemos uso das predições do modelo numa tarefa derivada, um teste de associação de palavras, como segundo critério para obtermos indícios de validade do ajuste do modelo ao estudo de caso. A escolha do teste de associação de palavras para validar o modelo não se dá por acaso. No contexto da Psicanálise, o problema de evidenciar a existência de relações entre as associações

livres e os complexos inconscientes foi endereçado pela aplicação do método de associação de palavras, partindo dos estudos empreendidos por Jung, e seus resultados foram tomados por Freud como comprovação experimental da técnica de associação livre, conforme discutimos na Subseção 1.3.3. Mas a relação entre o teste de associação de palavras e as representações semânticas não interessa apenas à Psicanálise. A facilidade de aplicação e a concretude dos testes de associação de palavra permitem utilizá-los como um crivo para avaliar diferentes hipóteses sobre o processamento cognitivo de informações semânticas. Griffiths et al. (2007), por exemplo, comparam a LDA com outros modelos baseados na representação semântica de espaço vetorial, avaliando o desempenho dos modelos numa tarefa de predição dos resultados de associações de palavras – e concluem em favor dos modelos de tópicos por capturar mais adequadamente os padrões associativos. Utilizar testes de associação de palavra para validar o modelo permite confluir tanto a importância desses testes na história da Psicanálise quanto sua aplicação contemporânea para validar modelos de representação semântica, um dos contextos de aplicação dos modelos de tópicos na Psicologia Cognitiva.

#### 4.6.1 Modelando o teste de associação de palavras

Durante a realização dos atendimentos, foram aplicadas cinco baterias de associações de palavra em diferentes sessões, cada uma contendo 20 palavras-estímulo. Mesmo aplicando simultaneamente várias palavras, consideraremos cada associação como independente de todas as outras dentro de uma mesma sessão. Ou seja, dada a sessão na qual foi aplicada a bateria, cada par estímulo-resposta é considerado independente de todos os outros. Apesar de esse pressuposto ser pouco verossímil – o analisando certamente é influenciado pelas palavras-estímulo e palavras-resposta anteriores – ele facilita a computação das quantidades de interesse para validação do modelo, descritas logo abaixo.

No modelo que propomos para as associações livres, as representações-meta são amostradas dos complexos inconscientes proporcionalmente à força do investimento libidinal num dado momento. No teste de associação de palavras essa liberdade é desfeita pela utilização de uma palavra-estímulo que irá influenciar as representações-meta, as quais, por sua vez, guiarão as possibilidades de resposta. Dado um determinado investimento libidinal sobre os complexos,

a palavra-estímulo *excita* justamente aqueles complexos mais fortemente associados à palavra. Em termos probabilísticos, a partir de uma distribuição sobre tópicos  $p(z)$ , a ocorrência de uma palavra  $w_1$  irá atualizá-la para a distribuição condicional  $p(z | w_1)$ . Será essa distribuição condicional que irá determinar quais termos terão maior ou menor probabilidade de serem associados, de acordo com a distribuição marginal  $\sum_{i=1}^K p(\vec{w} | z_i)p(z_i | w_1)$ . Essa intuição da relação entre os modelos de tópicos e as associações de palavras já havia sido proposta por Griffiths et al. (2007) no contexto da avaliação da probabilidade de associações de uma palavra específica, conforme a Equação 4.4.

$$p(w_2 | w_1) = \sum_{i=1}^K p(w_2 | z_i)p(z_i | w_1) \quad (4.4)$$

A definição de uma palavra específica como resposta,  $w_2$ , permite avaliar quão provável é uma palavra de fato associada sob o modelo ajustado. A computação da distribuição  $p(w_2 | w_1)$  não é simples e exige deduzir algumas quantidades a partir das definições do nosso modelo. A primeira dificuldade é definir a distribuição marginal sobre tópicos,  $p(z)$ , uma vez que o modelo ajustado só nos fornece a distribuição de tópicos para documentos específicos,  $\vec{\theta}_d$ . Para obtermos a distribuição desejada, é necessário integrar com relações a todos os valores possíveis de  $\theta$ .

$$p(z) = \int_{\Theta} p(\vec{\theta})p(z | \vec{\theta})d\vec{\theta} \quad (4.5)$$

A integral pode ser facilmente resolvida uma vez que percebemos que o termo  $p(z | \vec{\theta})$  é simplesmente igual a  $\vec{\theta}$ , o que reduz a integral à fórmula do valor esperado de  $\vec{\theta}$ ,  $E[\vec{\theta}]$ .

$$\begin{aligned} p(z) &= \int_{\Theta} p(\vec{\theta})p(z | \vec{\theta})d\vec{\theta} \\ &= \int_{\Theta} p(\vec{\theta})\vec{\theta}d\vec{\theta} \\ &= E[\vec{\theta}] \end{aligned} \quad (4.6)$$

Infelizmente, o modelo ajustado pela função **stm** não nos fornece esse valor diretamente. O que temos em seu lugar é o valor esperado  $\vec{\mu}$  da variável  $\vec{\eta}$ , ou seja, a proporção de tópicos *anterior* à sua transformação pela função logística. A primeira intuição seria

trabalhar apenas com a transformação logística de  $\vec{\mu}$ , mas o resultado obtido não está correto: o valor esperado da transformação logística de  $\vec{\eta}$  não corresponde à simples transformação de  $\vec{\mu}$  porque a função logística é não-linear. Para complicar ainda mais a situação, não existe uma solução analítica em forma fechada para computar  $E[\text{logit}^{-1}(\vec{\eta})]$ . A solução que encontramos é utilizar integração numérica:

1. Considerando o valor esperado  $\vec{\mu}_s$  relativo a uma sessão específica  $s$  e a matriz de covariância  $\Sigma$ , geral para todo modelo. . .
2.  $N = 5.000$  amostras aleatórias  $\vec{\eta}_s$  são sorteadas da distribuição normal multivariada  $\text{MultiN}(\vec{\mu}_s, \Sigma)$ ;
3. As amostras são transformadas pela função logística,  $\vec{\theta}_s = \exp(\vec{\eta}_s) / \sum \exp(\vec{\eta}_s)$ ;
4. O valor esperado  $E[\vec{\theta}]$  é aproximado pela média aritmética  $\frac{1}{N} \sum_{i=1}^N \vec{\theta}_s$ .

Com cinco mil simulações, o erro de aproximação em função da integração numérica é irrelevante, e conseguimos obter o valor de  $p(z)$ . A distribuição condicional  $p(z \mid w_1)$  é facilmente obtida por meio da regra de Bayes e dos parâmetros estimados pelo modelo, conforme a Equação 4.7.

$$\begin{aligned} p(z \mid w_1) &= \frac{p(z)p(w_1 \mid z)}{\sum_{i=1}^K p(z_i)p(w_1 \mid z_i)} \\ &= \frac{E[\vec{\theta}]\vec{\beta}^{(w_1)}}{\sum_{i=1}^K E[\vec{\theta}]\vec{\beta}_{z_i}^{(w_1)}} \end{aligned} \quad (4.7)$$

A distribuição *a priori*  $p(z)$  foi definida logo acima na Equação 4.6 como sendo igual a  $E[\vec{\theta}]$ . A distribuição condicional  $p(w_1 \mid z)$  é simplesmente a probabilidade da palavra  $w_1$  no tópico  $z$ , valor que pode ser recuperado do parâmetro  $\vec{\beta}$ , presente no ajuste do modelo. Com isso, temos todos os elementos necessários para calcular a Equação 4.4:

$$\begin{aligned} p(w_2 \mid w_1) &= \sum_{i=1}^K p(w_2 \mid z_i)p(z_i \mid w_1) \\ &= \sum_{i=1}^K \vec{\beta}_{z_i}^{(w_2)} \frac{E[\vec{\theta}]\vec{\beta}^{(w_1)}}{\sum_{i=1}^K E[\vec{\theta}]\vec{\beta}_{z_i}^{(w_1)}} \end{aligned} \quad (4.8)$$

Essa equação permite-nos calcular a probabilidade da palavra-resposta associada pelo analisando,  $w_2$ , condicionada à palavra-estímulo  $w_1$  no contexto do modelo de tópicos que ajustamos ao *corpus*. Mas isso, por si só, não é suficiente. É ingênuo acreditar que o modelo poderia prever exatamente a palavra associada como a de maior probabilidade, pois, como veremos nos resultados, o próprio analisando variou suas respostas para uma mesma palavra-estímulo, o que deixa claro que o processo não é determinístico, mas estocástico. Na ausência de um modelo concorrente adequado para avaliar o desempenho de nosso modelo de tópicos, utilizaremos a distribuição marginal dos termos do vocabulário. Apesar de pouco verossímil, a distribuição marginal pode servir de crivo para avaliar a seguinte situação: se o modelo de tópico não captura nenhuma estrutura latente nas associações livres, suas predições não seriam muito diferentes das frequência observadas para os termos. Essa mesma situação pode ser invertida no sentido da relação entre as associações livres e teste de associação de palavras: se não há nenhuma relação entre as palavras associadas e os complexos inconscientes, o analisando estaria apenas selecionando palavras aleatoriamente de seu vocabulário, com probabilidade proporcional à frequência observada dos termos.

### 4.6.2 Hipóteses

Com base na descrição feita sobre os testes de associação de palavra, as relações esperadas entre os complexos inconscientes e as associações livres, e tomando a distribuição marginal dos termos como linha de base para comparação, propomos duas hipóteses para testar nosso modelo escolhido ( $MT$  é o índice designando a distribuição de probabilidade sob o modelo de tópicos;  $MM$  designa o modelo marginal).

$$H_1 : \frac{p_{MT}(w_2 \mid w_1)}{p_{MM}(w_2)} > 1 \quad (4.9)$$

A primeira hipótese,  $H_1$ , prediz que a razão entre a probabilidade condicional da palavra associada, dada a palavra-estímulo, predita pelo modelo de tópicos,  $p_{MT}(w_2 \mid w_1)$ , e a probabilidade marginal da palavra associada,  $p_{MM}(w_2)$ , será maior do que um, pelo menos em termos médios. Isso implica que a palavra associada será mais provável sob o modelo de tópicos do que no modelo baseado na distribuição marginal das palavras. Exagerando um pouco



a confiança em nossos pressupostos, ao considerar cada par associativo como independente dos outros, podemos utilizar a produto das razões de todos os pares – ou a soma de seu logaritmo – como um teste de razão de verossimilhança dos dois modelos. Obviamente não é possível considerar esse procedimento um teste estatístico pela ausência de uma distribuição amostral de referência, mas ele permite dar uma dimensão da probabilidade relativa de cada modelo em função dos dados de teste, ou, mais especificamente, uma medida da *evidência* em favor do modelo de tópicos.

$$H_2 : \text{ranque}(p_{MT}(w_2 \mid w_1)) < \text{ranque}(p_{MM}(w_2)) \quad (4.10)$$

A segunda hipótese,  $H_2$ , prediz que o ranque da palavra associada – ou seja, sua posição na distribuição de probabilidade, ordenada de maneira decrescente em função da magnitude da probabilidade – será maior na distribuição condicional do modelo de tópicos,  $p_{MT}(w_2 \mid w_1)$ , do que na distribuição marginal,  $p_{MM}(w_2)$ , pelo menos em termos médios. Nesse caso, podemos propor um teste de hipótese rudimentar para avaliar quão surpreende são as ordens obtidas sob o modelo de tópicos, em comparação com o modelo marginal tomado como modelo nulo. Utilizando a distribuição do modelo marginal, sorteamos repetidamente amostras de palavras em quantidade igual ao total de pares do teste de associação e computamos o ranque mediano de cada amostra. Com isso, obtemos uma distribuição amostral sob a hipótese nula,  $H_0$ , de que o analisando está apenas escolhendo palavras ao acaso quando reponde a uma palavra-estímulo, e poderemos avaliar quão destoante são os ranques medianos do modelo de tópicos e do modelo marginal.

### 4.6.3 Resultados

Das 100 palavras-estímulo escolhidas do vocabulário do analisando, 18 delas foram utilizadas apenas uma vez, 14 foram utilizadas duas vezes, 14 foram utilizadas três vezes e três foram utilizadas quatro vezes, totalizando 49 termos. Os termos de estímulo mais frequentes foram “dissertação”, “Nina”<sup>1</sup> e “casa”. As respostas dadas pelo analisando foram compostas por 36 termos utilizados uma única vez, nove termos utilizados duas vezes, dez utilizados três vezes e três termos utilizados quatro, cinco e sete vezes cada um, num total

---

<sup>1</sup>O nome de uma namorada do analisando, conforme veremos no próximo capítulo.

de 58 termos. Entre as palavras-resposta, “Nina”, “mãe” e “tentar” foram os termos mais frequentes.

Dos 100 pares de estímulo-resposta, 56 são únicos, demonstrando a variação de respostas mesmo com a utilização de palavras-estímulo repetidas. 13 pares repetem-se duas vezes, e seis pares são repetidos três vezes, elencados na Tabela 4.6. Os pares mais comuns são, em sua maioria, palavras cuja relação semântica é de oposição ou complementaridade, com exceção dos pares “casa”-“mãe” e “dissertação”-“fim”, que estão relacionados a complexos específicos, como veremos no próximo capítulo.

**Tabela 4.6:** Pares associativos mais comuns

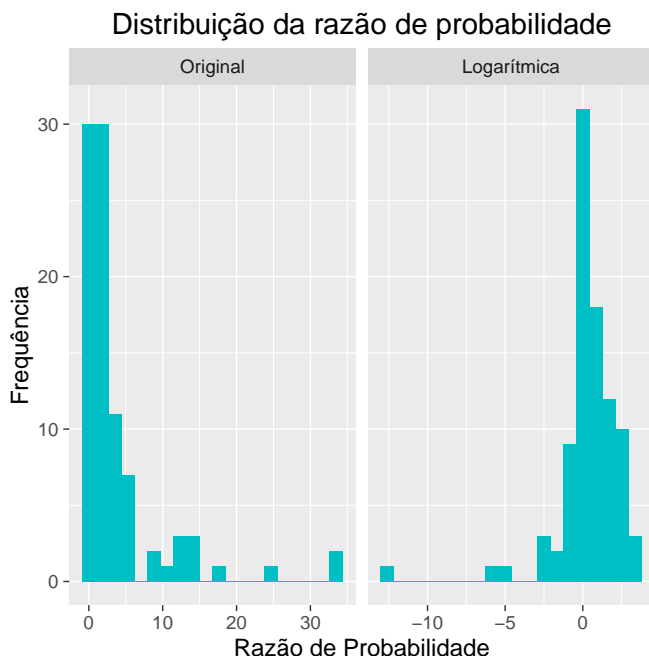
Estímulo	Resposta
Casa	Mãe
Querer	Poder
Dissertação	Fim
Diferente	Igual
Conseguir	Tentar
Vida	Morte

Antes da avaliação das hipóteses propostas acima, as palavras do teste de associação foram lematizadas e as *stopwords* presentes foram removidas. Esse procedimento é necessário para que o vocabulário do teste seja compatível com o vocabulário utilizado no ajuste do modelo. Após essa operação, restaram 91 pares de palavras.

#### 4.6.3.1 Razão entre probabilidades

Em 57 dos 91 pares de palavras (63%), a razão entre a probabilidade do modelo de tópicos e do modelo marginal foi superior a um. Apesar de o modelo proposto não dominar o modelo marginal de maneira absoluta, ele prevê as palavras associadas com maior probabilidade na maioria dos casos. Como pode ser conferido na Figura 4.11, a maioria dos pares de palavra tem razão de probabilidade próxima a um, estendendo-se até casos em que a probabilidade no modelo de tópicos é cerca de 33 vezes maior do que no modelo marginal. Os pares mais prováveis foram “Nina”-“Nina”, ou seja, quando ele repetiu a mesma palavra-estímulo referente a uma personagem importante em seus complexos. A escala original, porém, não faz jus às razões que favoreceram o modelo marginal, pois as

comprime entre zero e um. Por isso, acrescentamos também a visualização dos mesmos dados na escala logarítmica, que dá o mesmo valor absoluto para aumentos ou diminuições de mesma proporção, mudando apenas o sinal. O segundo histograma deixa evidente a presença de um par particularmente mal predito pelo modelo de tópicos: “vontade”-“potência”.



**Figura 4.11:** Distribuição da razão de probabilidade dos pares associativos na escala original (esquerda) e logarítmica (direita).

Na escala original, a média das razões de probabilidade é 3,8, e a mediana é 1,45. Como há valores extremos influenciando o cálculo do valor esperado, a mediana dá uma ideia mais robusta da tendência central: a probabilidade das palavras associadas é cerca de 45% maior no modelo de tópicos do que no modelo marginal. Na escala logarítmica, a assimetria da distribuição não é tão influente: sua média fica em 0,28 e a mediana em 0,37. Exponenciando a média do logaritmo (equivalente a calcular a média geométrica), obtemos um valor um pouco mais conservador, mais ainda assim favorecendo

o modelo de tópicos: as associações são 33% mais prováveis sob o modelo de tópicos.

Apesar da razão da probabilidade não ser consideravelmente grande em cada caso isolado, o fato de 57 de 91 pares terem apresentado um aumento com relação ao modelo marginal sugere uma considerável corroboração do modelo de tópicos. Essa corroboração pode ser medida em termos da *evidência* que os dados apresentam em favor do modelo de tópicos, conforme a sugestão de Jaynes (2003, p. 91):

$$e(H \mid DX) \triangleq 10 \log_{10} O(H \mid DX) \quad (4.11)$$

Ou seja, a evidência dos dados  $D$  com relação a uma hipótese  $H$  é medida como o logaritmo de base 10 das chances condicionais da hipótese em função dos dados. O termo  $O(H \mid DX)$  representa as chances posteriores do modelo de tópicos em relação ao modelo marginal, computada assumindo a independência entre os pares de palavras. :

$$O(H \mid DX) = \frac{p(H_{MT} \mid X)}{p(H_{MM} \mid X)} \quad (4.12)$$

$$= \frac{p(H_{MT}) \prod_{i=1}^N p(X_i \mid H_{MT})}{p(H_{MM}) \prod_{i=1}^N p(X_i \mid H_{MM})} \quad (4.13)$$

A multiplicação por um fator de 10 implica que a evidência é mensurada em decibéis, no sentido de expressar a intensidade da razão de dois valores – como comumente a medida é utilizada no caso da potência sonora. Os pares de palavras fornecem 112,6 dB de evidência em favor ao modelo de tópicos com relação ao modelo marginal, se não for considerada nenhuma diferença *a priori* entre a probabilidade dos modelos – uma evidência favorável *gritante*, com o perdão da palavra. Esse valor implica que, mesmo atribuindo uma probabilidade *a priori* extremamente pequena ao modelo de tópicos com relação ao modelo marginal – algo da ordem de  $10^{-8}$  – a evidência em favor do modelo de tópicos atualiza sua probabilidade para 0,999, o que torna o modelo de tópicos a escolha mais justificada entre os dois modelos mesmo se não estamos dispostos a confiar nele *a priori*.

#### 4.6.3.2 Ranque das palavras associadas

O ranque inverso das palavras associadas pelo analisando foi menor no modelo de tópicos do que no modelo marginal em 59 dos 91 pares (65%). Não por acaso, 55 dessas 59 palavras também tiveram a razão de probabilidade maior do que um, conforme os resultados acima, o que indica certa redundância das duas hipóteses propostas. Em média, a posição das palavras associadas na distribuição do modelo de tópicos estava 81 posições acima da distribuição do modelo marginal. A mediana é mais conservadora: metade das diferenças entre ranques está abaixo de 31 posições. O ranque mediano para as predições do modelo foi de 66, contra 78 do modelo marginal.

Um aspecto surpreendente da avaliação das posições sob o modelo de tópicos é que em 19 casos a palavra associada se encontrava entre as dez primeiras preditas pelo modelo. A Tabela 4.7 apresenta esses casos. Os primeiros são relativos aos pares de palavras nas quais a associação é uma repetição da palavra-estímulo. Em seguida, há pares relacionando palavras semanticamente próximas no sistema da língua, como “primeiro”-“segundo”, e pares peculiares à história do analisando, como “sítio”-“vô” e “vera”-“sevastopol”. A associação “mãe” ficou particularmente bem colocada frente aos estímulos “pai” e “casa” – o motivo da diferença de ordem para um mesmo par associativo se deve ao fato de eles terem ocorrido em sessões diferentes.

Os dados do modelo marginal também podem ser utilizados para testar uma hipótese nula de interesse: se as palavras-estímulos não têm um efeito sobre as associações do analisando, suas respostas seriam amostradas aleatoriamente da distribuição marginal dos termos, proporcionalmente à sua frequência de uso. Em outras palavras, essa hipótese nula sugere uma ausência de estrutura latente influenciando as respostas dadas. Sob essa hipótese nula, construímos uma distribuição amostral por meio de simulações: repetindo 5.000 vezes, amostramos um conjunto aleatório de 91 palavras com probabilidade igual à distribuição marginal dos termos e computamos a mediana de seu ranque inverso. Pela distribuição amostral simulada, esperaríamos que a mediana das ordens ficasse em torno de 131. Quão esperado é o resultado obtido pelo modelo de tópicos, uma mediana de 66? A probabilidade desse evento ou outro mais extremo sob a hipótese nula é de apenas 0,3%, sugerindo que a hipótese de ausência de estrutura pode ser rejeitada em favor de nossa hipótese alternativa.

**Tabela 4.7:** Ranque inverso das palavras associadas sob o modelo de tópicos e o modelo marginal

	Estímulo	Resposta	Ordem MT	Ordem MM
1	tudo	tudo	1	20
2	casa	casa	1	29
3	nina	nina	1	32
4	nina	nina	1	32
5	escrever	ler	1	44
6	escrever	ler	1	44
7	livro	ler	1	44
8	sítio	vô	1	60
9	sítio	vô	1	60
10	mãe	casa	2	29
11	primeiro	segundo	2	124
12	casa	mãe	3	22
13	casa	mãe	3	22
14	casa	mãe	3	22
15	pai	mãe	3	22
16	pai	mãe	5	22
17	ano	ano	7	38
18	vera	sevastopol	8	27
19	ruim	bom	10	10

#### 4.6.4 Sensibilidade dos resultados à escolha do modelo

Para avaliar se os resultados dos testes de hipótese são robustos à escolha do modelo, repetimos as análises feitas com o modelo de 40 e de 60 tópicos, nos dois extremos da faixa julgada ideal para o *corpus*. No modelo com 40 tópicos, 53 dos 91 pares de palavras (58%) possuem a razão de probabilidade superior a um, enquanto no modelo com 60 tópicos essa proporção é de 52 para 91 (57%). A diferença com relação ao modelo escolhido é pequena, o que ajuda a corroborar os resultados obtidos.

No caso da hipótese sobre o ranque inverso das palavras associadas, o modelo com 40 tópicos apresentou ranques mais baixos em 56 dos 91 pares (62%). Para o modelo com 60 tópicos, essa proporção ficou em 54 de 91 (59%). Novamente, há pouca diferença com relação às proporções apresentadas acima para o modelo escolhido, o que nos permite ter mais segurança nos resultados obtidos.

### 4.6.5 Considerações

Uma primeira conclusão importante é que o modelo de tópicos não se mostrou absolutamente melhor do que o modelo marginal com relação às predições de  $H_1$  e  $H_2$ . Em outras palavras, as predições feitas pelas duas hipóteses não se mostraram verdadeiras para todos os pares de palavras-estímulo e palavras-resposta. Mesmo assim, o modelo proposto fez predições mais adequadas para a maioria dos pares de palavras. Com isso, reconhecemos que o modelo teve um desempenho aceitável, mas ainda é necessário avaliar melhor em quais aspectos o modelo falhou: se na representação proposta pelo modelo formal; no algoritmo de ajuste do modelo; na falta de informações no *corpus*; ou na própria relação entre associações livres e testes de associação de palavra. Nossa suspeita recai sobre a quantidade de informações presentes no *corpus*, talvez ainda muito pequeno para estimar os parâmetros com precisão, e sob os procedimentos de estimação, que são afetados pela multimodalidade da distribuição posterior.

Por outro lado, a superioridade absoluta do modelo de tópicos sobre o modelo marginal, apesar de altamente desejável e com maior força de corroboração das propostas da tese, é um objetivo excessivamente exigente para um modelo probabilístico em uma área ainda incipiente. Mesmo em situações em que um problema é reconhecidamente bem predito por um modelo, dificilmente obtemos um desempenho superior absoluto ao compararmos dois modelos, em função da natureza estocástica dos dados e a presença de ruídos que podem afetar as predições. No nosso caso, aceitamos diminuir o poder corroborativo dos resultado em favor de uma predição um pouco mais flexível, mas ainda assim importante: o modelo de tópicos deve ser melhor que o modelo marginal em média.

Sob essa condição, os resultados são consideravelmente mais favoráveis. Com relação à  $H_1$ , vimos que o modelo tende a prever as palavras associadas com maior probabilidade na maioria dos casos, utilizando apenas as informações que o modelo aprendeu a partir do *corpus* de associações livres. Com relação à  $H_2$ , obtivemos o resultado de que a ordem das palavras associadas é menor no modelo de tópicos, chegando inclusive a sugeri-las entre as mais prováveis. Esses resultados, em conjunto com o ótimo desempenho dos parâmetros do modelos na predição de critérios externos, a alta verossimilhança dos dados sob o modelo de tópicos e a rejeição da hipótese nula de ausência de estrutura no testes de associação de palavra, sugerem que podemos confiar nas inferências realizadas com

considerável segurança, apesar da evidência não ser completamente conclusiva em favor do modelo proposto, exigindo a revisão de seus pontos fracos.



## Capítulo 5

### O caso Ivan Ivanovich

A validação do modelo de tópicos que ajustamos ao *corpus* nos permitiu desenvolver alguma confiança com relação às inferências que podemos fazer. Ainda que tenhamos cedido na severidade de nossas hipóteses e, conseqüentemente, na força de evidência que as análises empreendidas implicam, estamos razoavelmente seguros de que o modelo de tópicos captura aspectos relevantes das associações livres e que o ajuste escolhido não está completamente inadequado. Com isso, podemos passar à apresentação do estudo de caso a partir dos resultados encontrados pelo modelo.

Não seguimos, contudo, nem o formato freudiano de exposição de caso, baseado na construção de uma longa narrativa com detalhes intrincados das associações e inferências das relações entre eventos da história do sujeito e seus sintomas, nem a apresentação de vinhetas selecionadas para apresentar evidências duvidosas da prática clínica com relação a algum constructo teórico. O caso é apresentado seguindo a tese central deste trabalho: o modelo de tópicos é uma representação formal aceitável do processo de livre associação e, por isso, as inferências do modelo podem ser utilizadas como sumário dos principais elementos do caso. Focamo-nos em três aspectos principais:

1. O conteúdo e relações entre complexos:
  - a) Identificando os principais termos dos complexos por meio do parâmetro  $\beta$ ;
  - b) Analisando as principais correlações entre complexos estimados por  $\Sigma$ ;
  - c) Apresentando excertos dos turnos de fala mais representativos, de acordo com  $\theta$ .
2. A relação com o analista:

- a) Identificando complexos ligados às intervenções do analista, tal como estimado por  $\gamma^{inter}$ .

### 3. A dinâmica libidinal durante o tratamento:

- a) Verificando a flutuação dos complexos mais importantes em função do tempo, de acordo com  $\gamma^{sessão}$ .

Como o caso é apresentado a partir das inferências do modelo, evitamos construir qualquer tipo de explicação para os resultados fora do que é estabelecido no modelo de tópicos proposto. Essa postura de forma alguma é propositiva: a utilização dos sumários do modelo não deve servir, em nenhum momento, de substituto à aplicação crítica da Psicanálise para interpretação dos resultados. Nosso objetivo, porém, é mostrar que os resultados obtidos são expressivos por si só, podendo de fato estabelecer um vínculo mais claro entre os dados clínicos e como eles informam a teoria psicanalítica.

## 5.1 Apresentação

Para que alguns dos termos presentes nos resultados sejam mais prontamente compreensíveis, é necessário introduzir algumas referências básicas sobre o analisando cujo caso é apresentado aqui. Focamos nas associações apresentadas durante a primeira entrevista, pois, como veremos e como é de conhecimento comum em Psicanálise<sup>1</sup>, as associações e ações sintomáticas então manifestas costumam ser reveladoras dos complexos que dominam a neurose do analisando.

Ivan Ivanovitch<sup>2</sup> é um jovem de 26 anos, já graduado e cursando pós-graduação durante a época do tratamento. Natural da cidade de Sebastopol, estava morando, até o fim do tratamento, na capital, Moscou, mas planejava voltar à cidade natal. Ivan já iniciou processos psicoterápicos em duas ocasiões, tanto em sua cidade natal quanto no serviço psicológico onde foi atendido para a presente pesquisa, mas os detalhes sobre esses tratamentos prévios só apareceram lentamente nas sessões posteriores.

<sup>1</sup>Conferir, p.e., a discussão no artigo *Sobre o início do tratamento*: “Os primeiros sintomas ou ações fortuitas do paciente, tal como sua primeira resistência, podem possuir interesse especial e revelar um complexo que dirige sua neurose” (Freud, 1984h, p. 181).

<sup>2</sup>Conforme já notado acima, todos os nomes próprios foram anonimizados e detalhes que possam facilitar a identificação do analisando foram editados ou retirados por completo.

Instruído sobre a regra da associação livre desde o primeiro momento, os tópicos abordados na primeira entrevista já situam os núcleos temáticos que retornaram durante todo o tratamento. Em sua primeira associação, relata que buscou terapia anteriormente e agora, novamente, para tentar sair do furacão que se tornou sua vida e tentar entender melhor o que estava acontecendo com ele.

Seguindo em suas associações, Ivan desata o nó condensado sob o termo “furacão”, e associa, *en passant*, sobre vários temas que se mostraram cruciais ao longo de seu tratamento. O primeiro tema associado diz respeito ao término da dissertação do mestrado ou, em contrapartida, seu abandono. Afirma estar passando por intensa dificuldade com relação à escrita e não confia na qualidade do que está fazendo. Como não consegue terminar o trabalho, oscila entre jogar tudo para o alto e esquecer a pós-graduação ou retomar o que já fez com cuidado para poder finalizar a dissertação de uma vez por todas.

Essa temática logo cede lugar aos planos de retorno à cidade natal de Sevastopol, em conflito com o interesse em continuar morando na capital. O sentimento de ambiguidade entre a permanência ou não em Moscou aparece vinculado diretamente a outros dois temas, mesmo que mais tarde venha se revelar como uma questão relativamente independente. A ideia de retorno a Sevastopol não lhe é atrativa, em parte, porque ficará muito próximo da mãe, com quem diz ter uma relação muito difícil. A maneira como a mãe o trata, exigindo dele se mostrar sempre de bem-estar, lhe causa um asco profundo, a ponto de desenvolver uma intensa raiva contra ela. Durante um breve retorno a Sevastopol, no ano anterior ao início do tratamento, a convivência constante com a mãe fez com que essa raiva chegasse a um extremo insuportável, idealizando de diversas maneiras a morte da mãe. O mal-estar causado por essas ideias, que Ivan considera completamente estranhas a seu modo de se relacionar com as pessoas, forçou-o a retornar a Moscou.

O segundo tema relacionado à mudança de Moscou para a cidade natal diz respeito ao fim do relacionamento explosivo que teve com uma amiga da universidade, Nina. Acreditava que, se se afastasse dos contextos que lhe traziam à memória as lembranças do relacionamento, poderia finalmente a esquecer. O desejo de se manter afastado de Nina é tão grande que chega a ter ideias suicidas, gastando horas em pensamentos sobre sua própria morte. Essas ideias basculam no sentido de tornar a própria Nina seu alvo, imaginando a morte dela de maneira violenta, mas geralmente

acidental. Ao mesmo tempo, Ivan insiste que as recordações de seu relacionamento com Nina e a dissertação inacabada são as únicas coisas que lhe restaram de um momento que considera especial de sua vida.

Uma das marcas da passagem para o atual momento conturbado foi a morte de seu avô materno após uma tentativa de suicídio. Ivan morou a maior parte de sua graduação no sítio de seu avô, e nutria por ele, ao mesmo tempo, um sentimento terno e uma constante aversão, especialmente porque o fazia executar tarefas que julgava ser completamente desnecessárias. Após a mudança para Moscou para iniciar o mestrado, acabaram se distanciando, afora os encontros usuais nas datas comemorativas da família. Quando o avô foi internado em estado grave, Ivan não foi visitá-lo nenhuma vez, e sente-se culpado por causa disso. Sua culpa, na verdade, vai mais longe: sugere que já havia percebido sinais de que o avô poderia não estar bem e que poderia atentar contra sua própria vida. Sua intuição era que o avô, após perder a esposa em função de um câncer, não tinha mais ninguém com quem conversar.

A importância de ter amigos ou conhecidos com quem conversar forma um último núcleo temático, abordado de maneira superficial na primeira entrevista, mas que vai se mostrar um elemento crucial das fantasias de Ivan. Desde a adolescência, Ivan sente um intenso *Wanderlust*, um desejo de viajar sem rumo pelo mundo. Dentre os interesses nessas viagens está a possibilidade de conhecer novos lugares e, principalmente, novas pessoas com quem possa conversar livremente sobre o que pensa. O desejo ardente de encontrar almas gêmeas — que surge tanto no sentido lato, de amigos com interesses em comum, quanto no sentido estrito, na figura idealizada de uma mulher especial — é comparado à experiência que teve ao se mudar para a capital e encontrar pessoas parecidas consigo, e contraposto às amizades da terra natal, de pensamentos tão distintos que faziam Ivan preferir permanecer calado a dar voz a seus pensamentos. Influenciado pelas leituras de Thoreau, a fantasia de encontrar pessoas em suas viagens cede lugar ao seu oposto: morar isolado de todo convívio social, sustentando-se apenas daquilo que ele mesmo produz. Esse ponto da fantasia volta a se ancorar nas lembranças de seu avô, pois Ivan imagina o sítio ou seus arredores como o contexto geográfico perfeito para experienciar o esplêndido isolamento.

## 5.2 Principais complexos

Os vários temas abordados rapidamente e de maneira encadeada na primeira entrevista são facilmente reconhecíveis entre os complexos identificados pelo modelo de tópicos. Além deles, outros conjuntos de complexos importantes também surgem ao longo do tratamento. Enquanto alguns dos novos temas se mostram significativos e fortemente relacionados com os temas para os quais a primeira entrevista serviu de exposição preliminar, outros, não menos importantes, permitem compreender como as associações circulam por assuntos de significância marginal ou mesmo apresentando oposição à regra de associação livre.

Nesta seção, apresentamos as inferências do modelo a respeito dos complexos a partir de três características de interesse. A primeira, possivelmente a mais relevante, é o conteúdo do complexo, caracterizado pelos 30 termos com maior probabilidade. A escolha dos 30 primeiros termos é consideravelmente arbitrária, mas a justificamos porque metade dos tópicos do modelo possuam 50% de sua massa de probabilidade entre os quatro e 30 primeiros termos. Nos casos em que os termos com maior probabilidade sejam excessivamente comuns e dificultem a interpretação do conteúdo, utilizamos a estratégia FREX. Esse algoritmo é sugerido por Roberts, Stewart, e Airoidi (2016) para construção de rótulos descritivos, baseado na média harmônica entre uma medida de exclusividade e de frequência dos termos dentro do tópico. A intuição por trás do algoritmo é que palavras com alta probabilidade são representativas do conteúdo de um tópico, desde que elas também não apareçam com alta probabilidade em outros. Procuramos manter, preferencialmente, o uso exclusivo da probabilidade de cada termo, e deixamos claro quando a lista de termos utilizada foi calculada por meio do método FREX.

A segunda informação apresentada é a expectativa, ou valor de probabilidade esperado, do complexo em função de qual interlocutor enunciou o turno de fala. O objetivo da apresentação desta informação é identificar se o complexo surge com maior frequência nos turnos de fala do analisando, indicando associações genuínas que sofreram pouca influência das intervenções do analista; ou nos turnos de fala do analista, indicando complexos que foram identificados como alvo das intervenções. Expectâncias próximas para o analista e o analisando num mesmo complexo pode servir como um indicativo insipiente de possíveis contaminações das interven-

ções sobre as associações do analisando. O maior problema dessa interpretação, porém, é que não é possível separar de maneira única, com as informações do modelo, o que é efeito sugestivo e o que é foco das intervenções.

A terceira informação, por fim, é a variação da expectância<sup>3</sup> da probabilidade do complexo em função do número da sessão. Demos a essa flutuação o nome de *dinâmica libidinal* justamente porque ela expressa como a relevância de um determinado complexo flutua em função do tempo. Essa informação pode ser de especial interesse para avaliação do efeito terapêutico da análise, uma vez que permite ter uma ideia de como determinados complexos emergem como centro de gravitação das associações num dado momento da análise, indicando seu forte investimento libidinal, para depois perderem força e cederem lugar a outros complexos no decorrer do tratamento, o que pode sugerir um desfecho positivo. Independente da aceitabilidade da redução da força de um complexo como um desfecho desejável do processo de análise, a flutuação dos tópicos é, por si só, informativa da dinâmica das associações.

Como complemento, para facilitar a interpretação dos termos componentes dos complexos, apresentamos também os turnos de fala considerados mais representativos pelo modelo. Apesar de ser complementar, essa informação indica uma utilidade importante dos modelos de tópicos para a apresentação de informações de um caso clínico, pois permite relatar excertos importantes das transcrições a partir de critérios definidos pelo próprio modelo e, portanto, consideravelmente independente de qualquer viés de seleção por parte do pesquisador. Assim, além de dar ao leitor uma ideia dos contextos a partir dos quais o modelo inferiu as relações de coocorrência entre os termos, esta informação também permite apresentar excertos qualitativos sobre o caso.

---

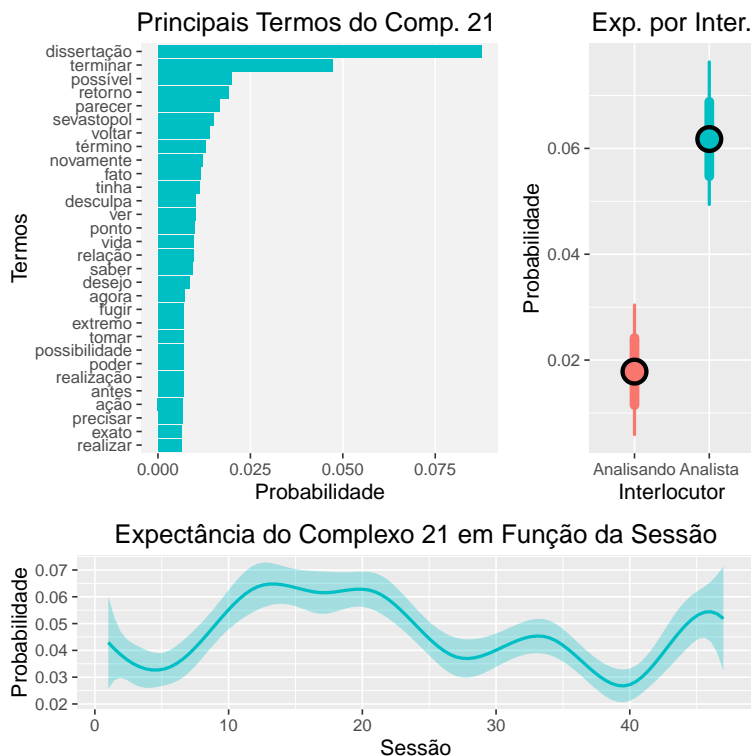
<sup>3</sup>Tecnicamente, o que as estimativas apresentadas representam não é exatamente a expectância do complexo com relação ao interlocutor ou à sessão, mas os coeficientes da equação de regressão definida na Subseção 3.3.1. Todavia, como os coeficientes de regressão indicam justamente diferenças no valor esperado em função de mudanças na variável preditora, e como só temos dois preditores incluídos no modelo, utilizamos diretamente o termo *expectância* como forma de facilitar a exposição dos resultados. Por isso, a presença de alguns valores negativos nas estimativas não são incoerentes – significam apenas que a expectância total, somando ambos os preditores, é reduzida pelo preditor cujo coeficiente é negativo.

### 5.2.1 Término da dissertação

Seguindo a ordem com que os temas emergiram na primeira sessão, iniciamos as análises dos complexos pelo conflito com o término da dissertação. O termo *dissertação* aparece com alta probabilidade principalmente no complexo de número 21 — alguns outros complexos nos quais o termo aparece também como importante são abordados abaixo. As principais inferências relativas a este complexo são apresentadas na Figura 5.1. O gráfico de barras, à esquerda, indica os 30 primeiros termos em ordem de magnitude de probabilidade; o gráfico de lagartas, à direita, apresenta a expectância, ou o valor esperado, do complexo de acordo com o interlocutor a quem pertence o turno de fala. O ponto, circulado em preto, é a média da distribuição posterior e serve de estimativa pontual do valor esperado. As linhas mais grossas representam o intervalo de um desvio padrão da distribuição posterior, análogo ao erro padrão. A linha mais fina indica o intervalo de probabilidade de 95%. A série temporal suavizada, apresentada na parte inferior, apresenta a flutuação do valor esperado do tópico em função do número da sessão. A linha é a média da distribuição posterior, e a banda mais clara representa o intervalo de um desvio padrão.

Os principais termos remetem ao término da dissertação, à fuga e busca de desculpas, e ao retorno à cidade natal. Curiosamente, essa maneira de abordar o complexo sobre a dissertação é mais provável nos turnos de fala do analista do que do analisando — cerca de três vezes mais provável. Portanto, a relação estabelecida entre o término da dissertação e o retorno a Sevastopol é um alvo das intervenções do analista, como pode ser conferido no turno de fala abaixo, que é documento com maior proporção do complexo 21.

Analista: Sem dúvida, a dissertação é utilizada como uma desculpa. Mas nesse caso é uma desculpa justificada, pois é terminando ela que terás opções que neste momento não tem. Inclusive com relação às tuas responsabilidades com o programa de pós-graduação. Se simplesmente abandonasse, teria que arcar com as consequências disso, e não seriam consequências positivas, em sua maioria. Ainda que também se visse livre. Mas mesmo que use a dissertação de desculpa, essa desculpa tem um prazo certo, ou mais ou menos certo, pelo menos não terás mais ela à tua disposição para usá-la como desculpas. Terás de tomar alguma decisão com relação



**Figura 5.1:** Gráfico das características do complexo 21. O primeiro gráfico, acima e à esquerda, apresenta os 30 primeiros termos e suas respectivas probabilidades no complexo. O gráfico à direita indica a expectância, ou valor esperado, do complexo em função ao interlocutor que enunciou o turno de fala. O ponto, em preto, indica a média da distribuição posterior, as linhas mais grossas indicam o intervalo correspondente a um desvio padrão, e as linhas mais finas indicam o intervalo de probabilidade de 95%. O gráfico abaixo apresenta a variação do valor esperado do tópico em função do número da sessão. A linha indica a média da distribuição posterior e a banda mais clara representa o intervalo correspondente a um desvio padrão.



a essas coisas que posterga em função da dissertação.  
(Sessão 33, Turno 32)

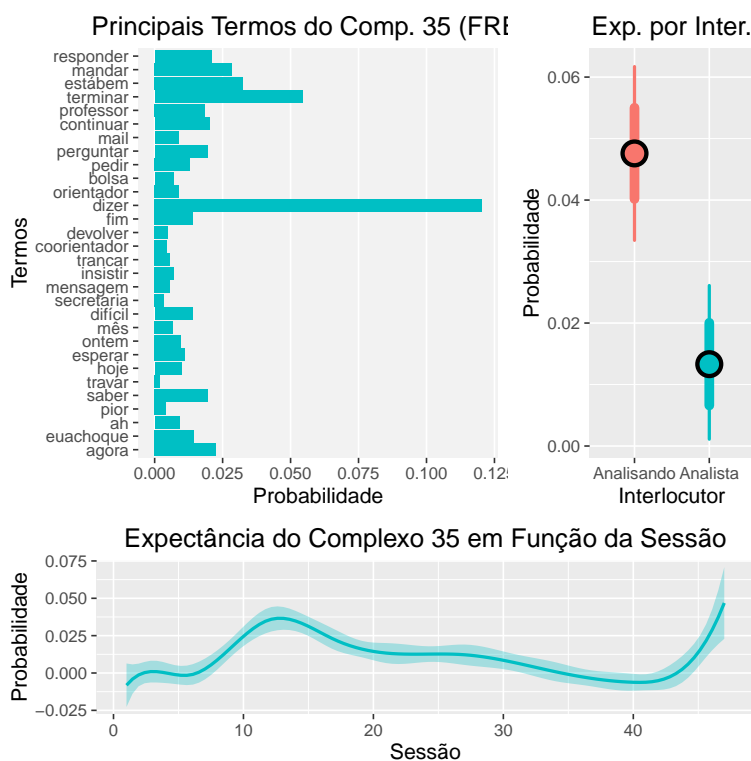
A interpretação estabelece uma justificativa para a procrastinação de Ivan em função do término da dissertação. Na medida em que não a termina, pode utilizá-la de desculpa para evitar tomar uma decisão com relação ao seu futuro. O analista também chama a atenção para o fato de que o término da dissertação tem um prazo bem delimitado e que seu desfecho positivo pode inclusive ampliar suas opções para decisões sobre o futuro. A prevalência deste complexo nos turnos de fala da analista, porém, permite pôr em questão a validade da interpretação: se o analisando aborda esse complexo com pouca frequência, seria justificado interpretar o adiamento do término da dissertação como busca de desculpas para evitar tomar uma decisão?

Por outro lado, os complexos nos quais o termo *dissertação* aparece com maior probabilidade nos turnos de fala do analisando versam sobre outros temas mais centrais, como o complexo 37, sobre as relações de Ivan com as mulheres, e o complexo 31, sobre o retorno à Sevastopol ou a permanência em Moscou – ambos são descritos em maiores detalhes abaixo, em suas respectivas seções. Essa característica das associações do analisando pode justificar a interpretação: o termo *dissertação* emerge marginalmente nos mesmos contextos que questões mais relevantes, o que justificaria seu caráter accidental se interpondo no caminho de decisões importantes.

Nesse caso em particular, a interpretação do analista está relacionada ao estabelecimento de uma relação entre diversos complexos do analisando, mas seu conteúdo é relativamente independente deles. A ausência de uma correlação entre a interpretação do analista e os tópicos motivadores nas associações do analisando nos parâmetros do modelo é esperada, uma vez que eles de fato não coocorrem nos mesmos contextos. Como veremos a seguir, nem sempre esse é o caso: há complexos que se apresentam com frequência parecida nas falas do analisando e do analista. O maior problema é interpretar exatamente o que isso significa, pois, afinal, pode representar tanto a realização de intervenções bem pautadas nas associações do analisando quanto a sugestão de temas pelo analista e a aceitação da sugestão por parte do analisando, que passa a associar guiado pela interpretação.

Mas há outro complexo que também versa sobre o término da dissertação como tema central, porém, o termo *dissertação* não é o mais relevante. O complexo 35, apresentado na Figura 5.2 utili-

zando o algoritmo FREX, versa principalmente sobre as comunicações com o orientador, as consequências do abandono do mestrado em termos da devolução dos valores da bolsa e, enfim, o fim e término do trabalho. Diferente do complexo ligado às intervenções do analista sobre o tema, o complexo 35 não aborda claramente o retorno à Sevastopol nem seu papel intermediário com relação a outros temas da vida de Ivan, e é cerca de cinco vezes mais prováveis nos turnos de fala do analisando.



**Figura 5.2:** Gráfico das características do complexo 35.

O turno de fala mais representativo do complexo 35 é justamente uma fala de Ivan sobre sua hesitação com relação ao término abrupto do trabalho da pós-graduação. Neste excerto, a principal preocupação é a questão da devolução dos valores recebidos como bolsista, oscilando entre a firmeza de trancar o curso mesmo sob o

custo de devolver todo o dinheiro, ou continuar tentando trabalhar na dissertação sob a pífia influência de um funcionário da secretaria do curso.

Ivan: Mas eu pesquisei, e tem alguns casos que não precisa devolver o dinheiro da bolsa. Mas, sabe, esse nem é o pior. Ano passado eu estava... No fim do ano eu estava convicto de que eu ia devolver, de alguma forma. Eu fui na secretaria e falei que ia trancar e queria saber como seria feita a devolução, porque eu sabia que tinha que devolver. [...] Mesmo que tranque a bolsa agora durante o curso, tem que devolver o que já ganhou. [...] Então eu já estava com isso na cabeça. Eu fui na secretaria e um dos caras que atende lá disse, mas como? E eu vi ele hoje de manhã, meu deus. E ele, mas como, não desista, vai, continua, faz, termina. E ficou insistindo nisso, eu saí de lá sentindo muito mal, embora tenha dito que ia trancar. O mais difícil agora é chegar lá de novo e falar com meu orientador. Ah, está bem. E nos casos que não precisa devolver, tem acidente ou força maior. Eu não sei o que é força maior. E doença grave, num parágrafo único, eu li. (Sessão 14, Turno 69, edições feitas para preservar o anonimato)

Tanto o complexo 21, sobre as intervenções do analista ligadas ao término da dissertação, quanto o complexo 35, sobre as dificuldades para o término da dissertação nas associações de Ivan, aparecem com maior probabilidade entre as sessões 10 e 15 e após a sessão 40, conforme os gráficos de série temporal da proporção esperada dos tópicos em função do tempo. Essa coincidência temporal pode justificar as intervenções do analista, ainda que seu conteúdo condense os termos de outros complexos que também ocorrem nestes momentos, como o complexo 31, sobre o retorno a Sevastopol. Os parâmetros do modelo de fato indicam que os complexos 31 e 35 têm uma correlação positiva relativamente forte, ou seja, tendem a co-ocorrer num mesmo turno de fala, e que quanto maior sua presença num determinado turno de fala do analisando, maior a probabilidade do tópico 21 no turno de fala seguinte do analista. Como ambos os tópicos surgem com maior probabilidade nas associações do analisando, é possível que a interpretação do analista, capturada pelo complexo 21, reflita um entendimento intuitivo dessa relação descoberta automaticamente pelo modelo.

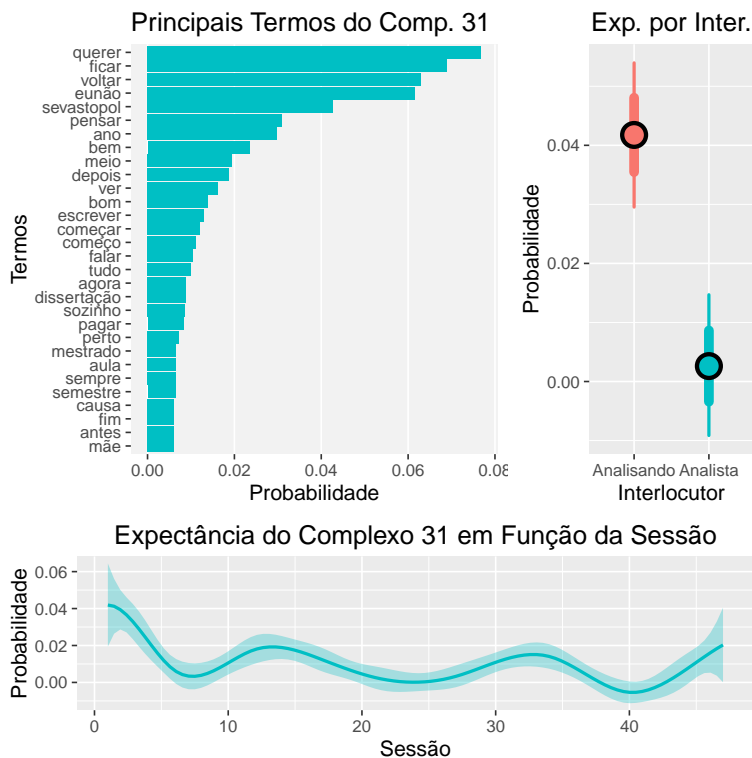
O modelo também identificou um complexo ligado ao tema da dissertação de Ivan e ao seu processo de escrita, mas as informações deste complexo poderiam levar à identificação do analisando<sup>4</sup>. Por este motivo, não o descrevemos aqui em detalhes. É importante ressaltar, porém, que esse tópico reúne termos relativos a outras atividades que Ivan gosta de realizar em seu cotidiano, indicando que o conflito com relação ao término da dissertação não provém de seu conteúdo. Pelo contrário, sua angústia se refere principalmente à falta de confiança de que sua escrita e os resultados apresentados tenham qualquer valor. A autodepreciação de suas opiniões, pensamentos e obras é um ponto importante nas associações de Ivan, e são abordados abaixo numa seção própria.

### 5.2.2 Retorno a Sevastopol

Para Ivan, o dilema do término da dissertação, além de constituir um núcleo temático próprio, aparece também de maneira marginal em outros complexos. Seguindo a intuição apresentada no complexo 21, relativo às intervenções do analista sobre o tema, seu aparecimento em conjunto com outros assuntos seria principalmente sob a forma de uma evitação de tomar uma decisão definitiva. Como vimos, um dos complexos em que o termo *dissertação* aparece versa sobre a questão do retorno à cidade natal de Sevastopol. As inferências do modelo sobre este complexo, de número 31, são apresentadas na Figura 5.3.

---

<sup>4</sup>A título de exemplo similar, Freud lamenta, no prefácio ao caso Dora, as restrições que é obrigado a impor sobre o material publicado do caso, que ele qualifica, inclusive, de “extraordinárias”. Sua preocupação, em particular, é que pessoas mal-intencionadas busquem desvendar os nomes reais que se escondem por trás do *roman à clef* que o estudo de caso pode vir a se tornar, prejudicando o paciente (Freud, 1989a).



**Figura 5.3:** Gráfico das características do complexo 31.

Não por acaso, os termos de maior probabilidade neste complexo possuem uma ambiguidade patente: *querer*, *ficar*, *voltar*, *eunão*. Se há uma afirmação de vontade, ela é seguida de perto por sua negação; e os verbos associados são antônimos, dado o contexto. Além do nome da cidade natal, o complexo ainda versa sobre a escrita da dissertação, a temporalidade do curso de mestrado e, ao fim dos 30 primeiros termos, também sobre a mãe, um dos motivos elencados na primeira entrevista para evitar o retorno. Para entender melhor a relação entre estes termos, apresentamos os cinco primeiros documentos indicados pelo parâmetro  $\theta$ , já que os primeiros documentos com maior proporção do complexo 31 são relativamente curtos e pouco informativos quando isolados.

Ivan: Sim, sim. Não, mas eu não conto muito com isso [vaga para concurso] porque o mestrado não é na minha

área, daí não ajuda. A dissertação, é só para me ver livre dela, mesmo. É que eu não li o Edital. (Sessão 33, Turno 37)

Ivan: Até tinha algumas vagas. Sempre tem para [minha área de conhecimento], mas eu não... Aí que está. Eu coloco, tem a dissertação, agora eu termino isso e depois eu penso em concurso, seja lá o que for. Agora eu não vou... (Sessão 33, Turno 35)

Ivan: Mas eu não queria ficar nessa guerrinha [com a mãe]. Eu queria... [...] Não me incomodar. (Sessão 16, Turno 47)

Ivan: E eu não queria fazer isso de novo com ela [Helena, uma ex-namorada]. Não seria legal repetir isso. [...] (Sessão 17, Turno 45)

Ivan: [Área de conhecimento da dissertação]. E estou terminando [a dissertação] agora. Ano passado eu voltei para Sevastopol para... Porque não precisava mais fazer disciplina do mestrado, eu estava pensando em voltar para lá. Estava tendo uns probleminhas aqui. E... Não foi muito legal. Porque eu voltei para a casa de minha mãe e eu não gosto dela. Na verdade para ela parece que está sempre tudo bem, mas para mim não está. Esse é um ponto. [...] Uma amiga minha, com quem eu estava ficando, que morava... Eu estava meio apegado, né. Queria me distanciar dela. Mas ao mesmo tempo que sabia que estava apegado, mas eu queria me distanciar porque ela não queria nada sério, e eu também não. Outra coisa é a dissertação, que eu não conseguia terminar. Essas três coisas me fizeram ficar bem mal, bem mal. Eu deitava no sofá da sala, olhava pela sacada... (Sessão 1, Turno 4)

Os dois primeiros excertos, peculiarmente, abordam diretamente a questão da dissertação e seu fim como critério para tomada de outras decisões importantes, como ficou mais explícito na seção anterior, mas não mencionam nem retorno a Sevastopol nem os termos mais prováveis identificados pelo complexo. A razão para esses dois excertos serem considerados mais representativos pelo modelo se deve principalmente ao seu tamanho. Como ambos os turnos de fala são relativamente curtos, formados por poucas palavras, não há infor-

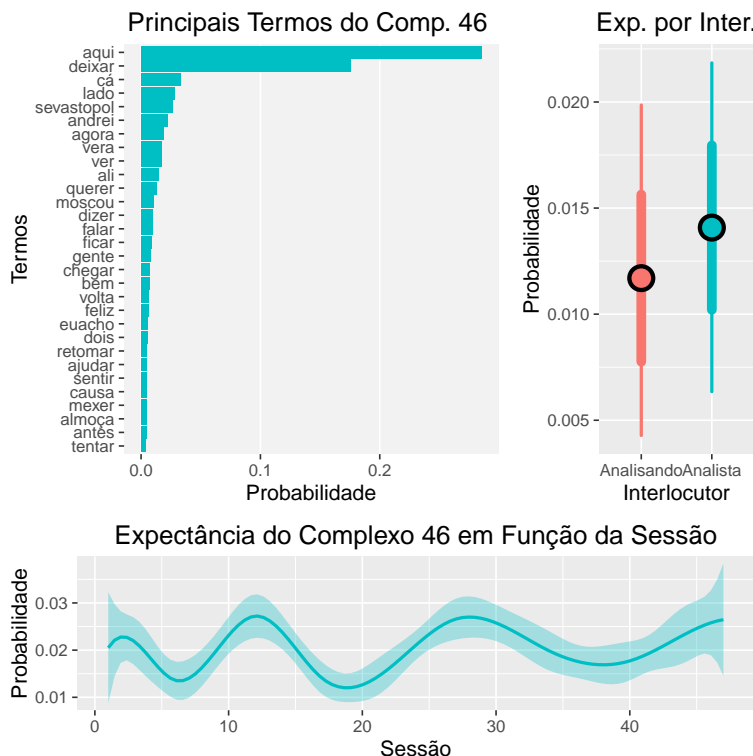
mação suficiente para identificar a presença de outros complexos, restando um deles com alta probabilidade.

Os dois excertos seguintes surgem, respectivamente, no contexto de associações a respeito da mãe e Helena, uma conterrânea com quem Ivan teve um breve caso amoroso após o rompimento com Nina. O último excerto, por fim, é proveniente da primeira entrevista e retoma aspectos que apresentamos na introdução do caso, como o retorno à Sevastopol e a relação com a mãe.

A variedade dos temas abordados nos turnos de fala representativos do complexo 31 indica uma maior heterogeneidade na composição deste tópico com relação ao que vimos acima. Essa heterogeneidade sugere que os diferentes termos estão fortemente associados entre si, e que o complexo não trata exclusivamente do retorno à terra natal, mas principalmente das relações pessoais implicadas por essa volta. Em particular, o fato de estes mesmos termos aparecerem como importantes em outros complexos indica a polissemia dos elementos componentes de um complexo. Apesar de qualquer termo ser um elemento único do vocabulário, seus contextos de coocorrência variam bastante e implicam em uma pluralidade de sentidos possíveis, cada um deles capturado por um complexo diferente.

O complexo 46, por outro lado, também trata do retorno à Sevastopol, mas mais precisamente da perspectiva de ir embora de Moscou e do afastamento dos amigos que fez na capital. Como pode ser conferido na Figura 5.4, os principais termos versam sobre *deixar*, *aqui*, *cá*, além do nome de vários amigos que Ivan fez desde sua mudança para Moscou.

Curiosamente, esse tópico é um pouco mais provável nos turnos de fala do analista, a julgar pela estimativa pontual do efeito do interlocutor, mas a incerteza em torno das estimativas impede diferenciar claramente um interlocutor preferencial para o uso do complexo. Os valores esperados em função da sessão para os complexos 31 e 46 indicam que eles não seguem exatamente o mesmo padrão temporal: as associações tendem a favorecer mais fortemente o conflito em retornar a Sevastopol e as relações estabelecidas por lá, enquanto a saída de Moscou e o distanciamento dos amigos aparecem com menor frequência. E, de fato, pelos parâmetros do modelo, não há nenhuma correlação clara entre os dois tópicos, o que sugere que, apesar do retorno implicar diretamente a despedida da capital, os complexos não formam dois lados de uma mesma moeda, mas são relativamente independentes entre si. A diferença entre os dois



**Figura 5.4:** Gráfico das características do complexo 46.

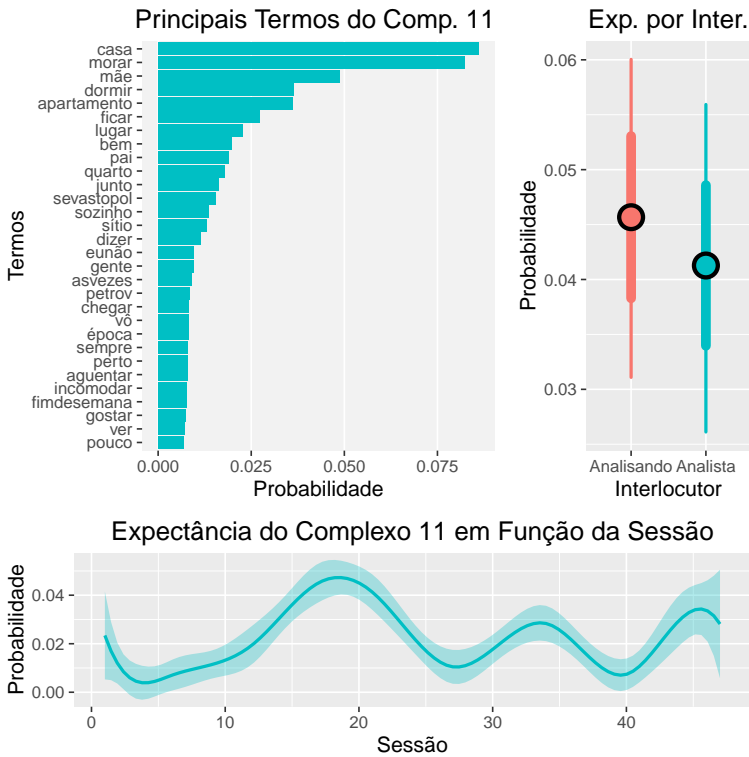
complexos mostra dois sentidos diferentes do termo *sevastopol*: é tanto um local para o qual retornar após o término da dissertação, onde a relação com a mãe e outros se tornará novamente presente, quanto o destino que implica o abandono de Moscou e dos amigos próximos com os quais Ivan tem dividido o cotidiano.

### 5.2.3 A mãe e o lar

O complexo 31 e o retorno à terra natal também são marcados pelo termo *mãe*. Apesar de não ser um termo tão importante neste contexto quanto *dissertação*, por exemplo, ainda assim indica a insistência com que a mãe de Ivan surge em suas associações. E, de fato, além de sua aparição substancial em outros complexos impor-



tantes, a ela é dedicado um complexo em particular, de número 11, cujas inferências são apresentadas na Figura 5.5. É possível notar que os principais termos são relacionados à moradia, como *casa*, *morar*, *apartamento*, *quarto*, *lugar*. O sítio do avô, também local de moradia de Ivan durante uma parte importante de sua vida, figura entre os termos junto com o nome de um importante amigo da cidade natal. Além da mãe, o termo *pai* aparece com grande probabilidade no complexo 11. O conjunto de termos dá mostras de outro aspecto da polissemia do termo *sevastopol*: agora, aproxima-se da ideia de lar, tanto no sentido de um lugar físico – seja a casa onde Ivan morou com os pais na infância, seja o apartamento para o qual se mudou após o divórcio, ou mesmo o sítio do avô – quanto no sentido interpessoal da relação familiar com o pai e a mãe, ou seus amigos.



**Figura 5.5:** Gráfico das características do complexo 11.

O complexo 11 aparece com mais ou menos a mesma probabilidade esperada tanto nos turnos de fala do analista quanto do analisando, com uma pequena preponderância nas associações de Ivan. De fato, apesar dos turnos de fala mais representativos pertencerem ao analisando, há vários turnos relevantes que pertencem ao analista, pertencentes principalmente às sessões 19 e 20, de maneira coerente com o que é predito pela série temporal. Apresentamos dois excertos das falas de Ivan, com base em sua representatividade do complexo 11, para discutir a variedades de temas específicos que este complexo abarca.

Ivan: Sim. [...] Eu não me vejo aqui [em Moscou] senão completamente infeliz. E eu vi no fim de semana que a Olga [amiga da universidade] faz falta. [...] E o Nikolai [amigo da universidade] não faz falta. E eu não me sinto mal por dizer isso. [...] Ele é que nem um amigo meu de Sevastopol, o Leon, que também não pára, é bem elétrico. Quando eu estava no sítio... [...] Na sexta eu ia a pé para a faculdade, dava uns dois quilômetros. Não era tanto, mas incomodava porque era morro acima. Eu começava a subir e chegava na casa dele, e a gente ia junto. E eu sentia saudade dele, porque, sei lá, eu ficava só com o vô no sítio, e ele era o primeiro amigo que eu via. Na época eu não conhecia o Petrov [amigo de Sevastopol] ainda, eu conheci quando já morava aqui. (Sessão 20, Turno 53)

Ivan: Eu não tenho muita vontade de querer ir morar com ele [pai], não. Ele mora meio longe da cidade, e isso talvez seja uma coisa ruim. Ele mora com a mulher dele, e eu acho que eu ia me sentir meio incomodando, talvez, porque ela fica o dia inteiro em casa. Não sei. Acho que eu não iria me sentir muito bem. Morando com a mãe teria a vantagem de... Ficar mais perto dos amigos. E... Eu me sentiria mais à vontade. [...] Porque é o lugar que eu consigo chamar de casa. [...] Eu fui para lá quando eu tinha uns doze anos, para o apartamento. Quando eles se separaram. E... Foi aí que... Sei lá, tornei o que eu sou agora, ou boa parte disso. (Sessão 12, Turno 21)

Analista: Mas parece aquela piada do cúmulo da revolta, morar sozinho e fugir de casa. Não é que não tens uma

casa, tens um lugar onde mora, por tua própria conta, junto com amigos e colegas. Não parece ser um lugar geograficamente localizado do qual queres fugir, mas de uma situação. (Sessão 19, Turno 18)

O primeiro turno trata, mais uma vez, do conflito entre a volta a Sevastopol e a permanência em Moscou, encarnado sob a figura de alguns amigos que Ivan tem nos dois lugares. Neste caso, o complexo aparece principalmente com relação ao local de moradia: o sítio, onde Ivan morava durante a graduação, e a casa de Leon, amigo que o acompanhava nas caminhadas para a faculdade. Leon é introduzido por analogia a um amigo da capital, Nikolai, por compartilharem a característica de serem extrovertidos e, apesar de serem bons amigos de Ivan, cansarem-no rapidamente com sua efusividade.

O excerto seguinte apresenta mais propriamente os termos de maior probabilidade no complexo 11. A casa do pai é apresentada como um lugar distante, tanto geográfica quanto emocionalmente. Ivan prefere permanecer, no momento, no apartamento onde mora a mãe e onde morou durante quase toda sua adolescência, tanto por poder ficar mais próximo de seus amigos de Sevastopol quanto por sentir que lá é seu lar, de fato. No contexto deste complexo, o termo *mãe* é aproximado da noção de lar.

A relação com o pai apresentada aqui é ambígua: ao mesmo tempo em que ele compartilha desta noção de lar que inclui a mãe, os excertos indicam que ele e Ivan não são próximos ou, pelo menos, que Ivan não se sente de todo confortável morando junto com o pai. O termo *dormir*, que aparece com alta probabilidade sob este complexo, pode lançar alguma luz sobre essa relação, se nos apoiarmos um pouco mais na teoria psicanalítica: a mãe de Ivan costumava dormir com ele durante a infância em função de episódios de terrores noturnos. Apesar de insistir em não ter tido muito interesse em pensar a respeito do processo de divórcio dos pais, focando nos possíveis benefícios que isso lhe trouxe, Ivan comenta que reflete, em alguns momentos, sobre como esses hábitos podem ter contribuído para a separação.

A intervenção do analista apresentada no terceiro excerto retifica uma queixa apresentada por Ivan. Comentando sobre um episódio da vida de uma amiga da faculdade, que fugiu de casa por um dia, reclama por não ter, no momento, uma casa de onde fugir, apesar de surgir em seu espírito um desejo intenso de fuga – cujos detalhes são abordados abaixo, no complexo sobre *Wanderlust*. A

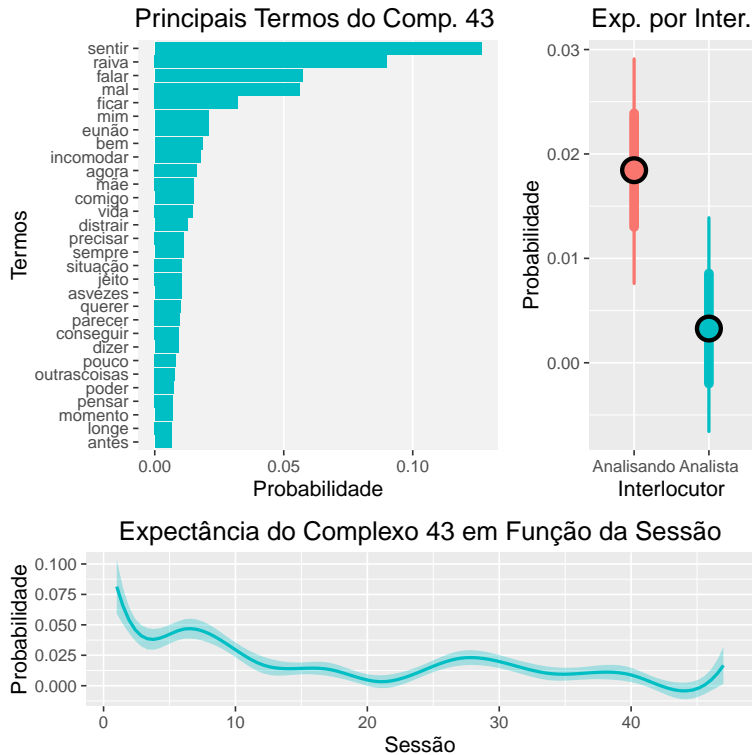
interpretação do analista situa a questão da fuga para seu contexto original: não se trata de fugir de um lugar geograficamente localizado – que por sinal Ivan já possui, a despeito de sua queixa – mas da situação colocada pelo dilema de retornar para o convívio com a mãe, em Sevastopol, ou permanecer afastado dela, em Moscou.

Podemos considerar, com alguma liberdade na interpretação do seu conteúdo, que o complexo 11 versa, pelo menos tangencialmente, sobre as relações edipianas de Ivan. A relação familiar com o pai e a mãe são colocados de maneira direta, sob a forma da grande proximidade da mãe no período da infância e um denegado sentimento de culpa com relação ao divórcio; mas também de maneira indireta, formulada na questão a respeito de seu verdadeiro lar, seja como moradia ou como cidade de domicílio. A proximidade do valor esperado deste tópico para analisando e analista sugere que esse fato não passou despercebido para o analista, que abordou os temas associados em suas intervenções.

#### 5.2.4 Raiva e mal-estar

Mas a familiaridade com que o termo *mãe* aparece neste contexto contrasta com seu aparecimento no complexo 43, que versa principalmente sobre o sentimento de raiva que acomete Ivan em diversas situações. Como pode ser conferido na Figura 5.6, os termos *mim*, *mãe*, *vida* sugerem possíveis destinos intercambiáveis da ira de Ivan. Em particular, o termo *incomodar* é usualmente utilizado para caracterizar as provocações que a mãe tipicamente lhe dirige quando ele está em Sevastopol, e que irritam Ivan profundamente, a ponto de ele não só pensar negativamente a respeito dela, mas também de falar com ela de maneira rude. Mas o sentimento de raiva não emerge de maneira isolada: o complexo também aborda, com alta probabilidade, os termos *ficar*, *mal*, *eunão* e *bem*. Em outras palavras, o modelo consegue capturar com precisão a passagem dos sentimentos de raiva aos sentimentos de mal-estar e culpa – *culpar*, por exemplo, está na 33<sup>a</sup> posição, próxima do fim da lista dos termos mais importantes listados.

A temática da raiva e seus destinos é cerca de quatro vezes mais provável nos turnos de fala do analisando do que do analista. A variação de sua ocorrência em função da sessão, como indicado no gráfico de série temporal, sugere um possível efeito positivo do tratamento. Nas primeiras sessões, até a sessão sete ou oito, a probabilidade esperada do complexo 43 era consideravelmente alta,



**Figura 5.6:** Gráfico das características do complexo 43.

especialmente se compararmos a escala do eixo *y* no Figura 5.6 com os gráficos anteriores. Ao longo das sessões, a prevalência deste complexo vai diminuindo paulatinamente, com pequenas flutuações ao fim do tratamento. O turno de fala mais representativo do complexo pertence, de fato, às primeiras sessões.

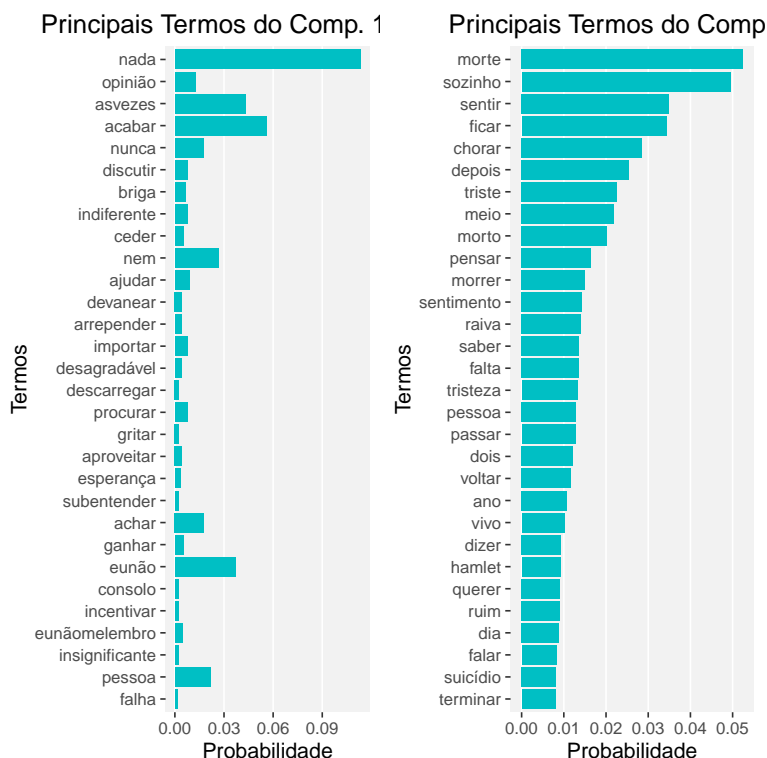
Ivan: Com a Nina sempre começou assim. Eu sempre soube que era assim. Eu sempre falava para ela que era assim, quando eu sentia raiva dela, quando a gente estava junto, ainda. Eu sentia raiva de mim, porque eu fiz tal coisa que não era muito razoável, e depois eu transferia essa raiva para ela porque eu não conseguia sentir raiva de mim muito tempo. Não me senti bem com isso. Aí era melhor sentir raiva de outra pessoa do que de mim

mesmo. E agora talvez seja isso. É ruim sentir raiva de mim. [...] O que eu vou fazer se eu sinto raiva de mim? [...] Ser ou não ser, né? (Sessão 5, Turno 43)

Apesar de não aparecer entre os termos mais prováveis, *nina* é também um termo com alta probabilidade no complexo 43, na 58ª posição. O excerto aborda o aspecto mais agressivo da relação com Nina, a colega de Ivan com quem ele teve uma relação conturbada e que o marcou profundamente. Nina costumava incomodá-lo, num sentido muito próximo daquele atribuído à mãe, fazendo-o sentir-se culpado por coisas que fizera e por coisas que ela lhe imputava. A reclamação constante sobre essas situações despertava uma profunda ira em Ivan, inicialmente dirigida contra ele próprio, e posteriormente dirigida contra ela. Nesta fala, em particular, o *não se sentir bem* se aplica à raiva direcionada contra si mesmo, mas o mal-estar também aparece pelo direcionamento da raiva contra outros, principalmente sob o termo *culpa*.

O termo *raiva* também aparece com alta probabilidade em outros dois complexos. No complexo 18, que possui forte correlação positiva com o complexo 43, aborda as estratégias utilizadas por Ivan para evitar entrar em conflito direto com outras pessoas. Essa estratégia passa principalmente por diminuir o valor de sua opinião ou simplesmente evitando afirmá-la quando ela pode provocar alguma discussão ou briga. Utilizando o critério FREX para identificar os elementos mais significativos, em função da presença de palavras genéricas como *nada*, *dizer* ou *às vezes*, os principais termos do complexo 18 são *opinião*, *discutir*, *briga*, *ceder* e *indiferente*, conforme o Figura 5.7. O turno de fala mais representativo deste tópico aborda um exemplo do uso desta estratégia de evitação na relação com Nina:

Ivan: Um exemplo, a gente começava a discutir sobre alguma tema e a gente tinha opiniões diferentes. Eu não sou muito de discutir. Eu já fui, quando eu era adolescente, até os vinte e poucos, vinte e um anos. Daí, depois, eu parei, eu perdi o ânimo de tentar fazer a outra pessoa mudar de opinião ou tentar passar uma ideia quando eu vejo que não tem por quê. Vai continuar do jeito que está. E é minha opinião também. E a gente chegava num ponto em que eu cansava daquela discussão. Então eu dizia, é assim. Ela [Nina] dizia que eu estava sendo indiferente, e eu estava. Mas eu não me importava e



**Figura 5.7:** Gráfico dos termos dos complexos 18 e 12.

eu acabava dizendo que eu não me importava mais com aquilo, com o que ela estava dizendo. E ela também não precisava se importar com o que eu estava dizendo porque eram coisas insignificantes. E eu dizia isso para ela. E ela dizia que não, é minha opinião, não é insignificante. Daí começavam as brigas. [...] Sem querer, porque eu estava querendo parar a discussão e acabava criando uma briga. Eu insistia, depois que já estava na briga eu insistia naquilo e discutia mais. Ela começava a gritar e eu ficava nervoso. E a briga se prolongava. [...] Talvez dessa forma. (Sessão 29, Turno 31)

O excerto revela, entretanto, que o resultado obtido por se mostrar indiferente ou dando pouco valor à sua opinião, pelo menos nas

discussões com Nina, nem sempre é próximo do esperado. Neste contexto, em particular, quando Ivan depreciava seu ponto de vista, Nina exigia dele que desse a consideração devida, e isso levava a novos conflitos. O complexo e o excerto dão alguma medida do caráter explosivo da relação entre Ivan e Nina, marcado por diversas brigas e discussões. Aqui, todavia, Nina é um dos destinos da ira de Ivan e alguém com quem deve aplicar sua tática de evitação de conflito; o papel de Nina como mulher e relação amorosa é tema de outro complexo, apresentado abaixo.

Antes de aprofundar a relação de Ivan com Nina e as mulheres, resta ainda abordar o complexo 12, cujo conteúdo já foi apresentado na Figura 5.7, e que também possui o termo *raiva* com alta probabilidade. Neste complexo, porém, o que se apresenta não é uma estratégia para evitar o surgimento do sentimento de raiva, mas o resultado extremo do direcionamento da ira contra si mesmo. Junto com o mal-estar despertado pelo sentimento de raiva contra si, Ivan também sente uma tristeza profunda, marcada pela solidão e acompanhada por ideias sobre a morte e o morrer. O turno de fala mais representativo do complexo 12 é um bom exemplo desse profundo pesar:

Ivan: Pois, porque eu estava lá quieto, meio morto. Eu acho, talvez. Então seria um jeito de me mostrar que eu estou vivo. Mas foi aquela explosão de raiva só, porque eu fiquei com raiva. Depois foi um misto de raiva e aquela meia morte, de novo. (Sessão 14, Turno 23)

Essa curta fala aparece no contexto de um conflito com a mãe, após discutir com ela as opções possíveis de moradia caso Ivan voltasse a Sebastopol. Sua mãe sugere que ele pode morar com ela, sem problemas, mas teria de pagar uma pensão próxima ao valor que ele gasta com aluguel em Moscou. A sugestão pretensiosa da mãe deixa-o extremamente irritado, a ponto de Ivan se fechar e permanecer calado. Em contrapartida, sua mãe passa a provocá-lo, até o momento em que ele explode, dando lugar à explicação descrita no excerto acima.

As ideias sobre a morte e o morrer passam por diversas modalidades nos pensamentos de Ivan. Muitas vezes são pensamentos que refletem a sensação de estar morto em vida, como no exemplo acima, preso em seus vários conflitos cujas resoluções são postergadas e que o deixa com a sensação de desamparo e desesperança. Em outros momentos, esses pensamentos cedem lugar a ideias suíci-



das propriamente ditas. Seus pensamentos perambulam por cenas de sua morte ou reflexões sobre procedimentos que poderiam tornar sua morte menos dolorosa. Mas dificilmente as associações se detêm na morte, simplesmente. Ivan entretém fantasias sobre os desfechos que sua morte poderia trazer para as pessoas que lhe são próximas, imaginando quem o visitaria no hospital ou em seu velório, ou como sua mãe ou sua irmã lidariam com o evento.

A relação entre a agressividade e o sentimento de culpa, bastante evidente nas associações de Ivan e capturada com precisão pelo modelo, é reconhecida na Psicanálise desde o momento em que Freud se viu levado a formular a ideia de uma pulsão de morte, ligada aos impulsos agressivos e originalmente dissociada das pulsões eróticas. Essa ligação seria a de transformação em função do recalque, pois, assim como “quando uma tendência pulsional sucumbe ao recalque, seus elementos libidinais se transformam em sintomas, seus componentes agressivos, em sentimento de culpa. Ainda que seja apenas aproximativamente correta, esta frase merece o nosso interesse” (Freud, 1989b, p. 163). No caso de Ivan, de acordo com as inferências feitas a partir do modelo de tópicos, a raiva é um dos componentes do complexo ligado à tristeza, pesar e mal-estar em geral, chegando ao extremo das ideias suicidas – um desfecho do retorno das pulsões agressivas sobre o próprio sujeito. Essa característica de pacientes obsessivos é enfatizada por Freud (1982) no famoso caso *O Homem dos Ratos*: de maneira geral, as ideias suicidas de seu paciente surgiam sempre quando ele era acometido de grande raiva contra alguém, que surgia concomitantemente com o sentimento de culpa em função dessas ideias agressivas.

### 5.2.5 Nina e as mulheres

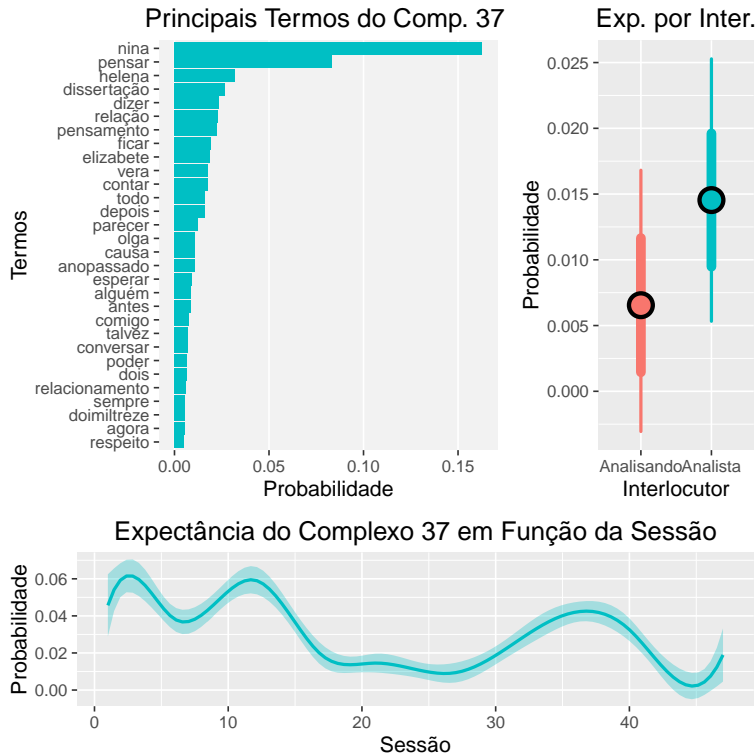
Vimos acima que Nina é um dos destinos da fúria de Ivan, principalmente por suas constantes provocações e acusações. Neste contexto, Nina é identificada com a mãe por uma característica comum, o que justifica seu agrupamento no mesmo complexo que trata do sentimento de raiva. Mas, assim como a mãe é, ao mesmo tempo, a referência do porto seguro ao qual Ivan fantasia retornar e o acosso que lhe desperta ira, tristeza e ideias mortíferas, Nina também oscila entre o papel odioso compartilhado com a mãe e o lugar de namorada, em conjunto com outras mulheres com quem Ivan teve um relacionamento íntimo.

Os principais termos do complexo 37, que aborda a relação de Ivan com as mulheres, são apresentados no Figura 5.8. O termo *nina* desponta como o de maior probabilidade, seguido por *pensar*, *helenas*, *dissertação*, *dizer*, *pensamento*, *ficar* e *elizabethes*. Falamos sobre Helena quando abordamos o complexo sobre o retorno a Sevastopol. Após o rompimento com Nina, Ivan retorna à terra natal no ano anterior ao início do tratamento e lá se envolve, por um curto período, com Helena. Desenvolve por ela um sentimento terno, mas não consegue deixar de pensar que o novo relacionamento não é mais do que uma distração, especialmente para evitar lembrar de Nina. Elizabethes, por outro lado, é uma colega com quem Ivan teve apenas alguns encontros. Sua significância provém pela forma como esses encontros se deram: no começo do relacionamento, antes de aprofundarem seus laços de compromissos, Nina não se mostrava disposta a manter Ivan como companheiro e, em suas próprias palavras, acabou oferecendo-o a uma amiga, Elizabethes. A aproximação entre Ivan e Elizabethes se deu num período em que ele e Nina decidiram não mais se encontrar, mas acabou despertando ciúmes em Nina e motivou diversas discussões entre os dois, até finalmente reatarem.

Apesar da relevância do tópico para a vida pessoal de Ivan, o complexo 37 também tem grande probabilidade nos turnos de fala do analista, conforme as estimativas do valor esperado em função do interlocutor. De fato, os turnos de fala mais representativos deste complexo são intervenções do analista, dos quais apresentamos o mais relevante, seguido pelo primeiro turno representativo enunciado pelo analisando.

Analista: Sim, o turbilhão de confusão do qual você falou na primeira sessão, inclusive. Da dificuldade que tinha nas relações sexuais com a Helena em função de estar com a Nina na cabeça. Mas por que, a cada pensamento que tens, tens que acrescentar a Nina ao final deles? (Sessão 3, Turno 43)

Ivan: Isso tem complicado bastante, esses pensamentos. Porque qualquer coisa que eu me imagine fazendo, eu imagino que estou sendo motivado a fazer aquilo por causa da Nina, de alguma forma. Tanto o concurso... Eu me empolguei bastante com o concurso, e depois eu pensava que estava empolgado só por causa da Nina. E daí eu comecei a ter essa ideia de sair por aí, e pensava que era só por causa da Nina. Isso, para qualquer coisa.

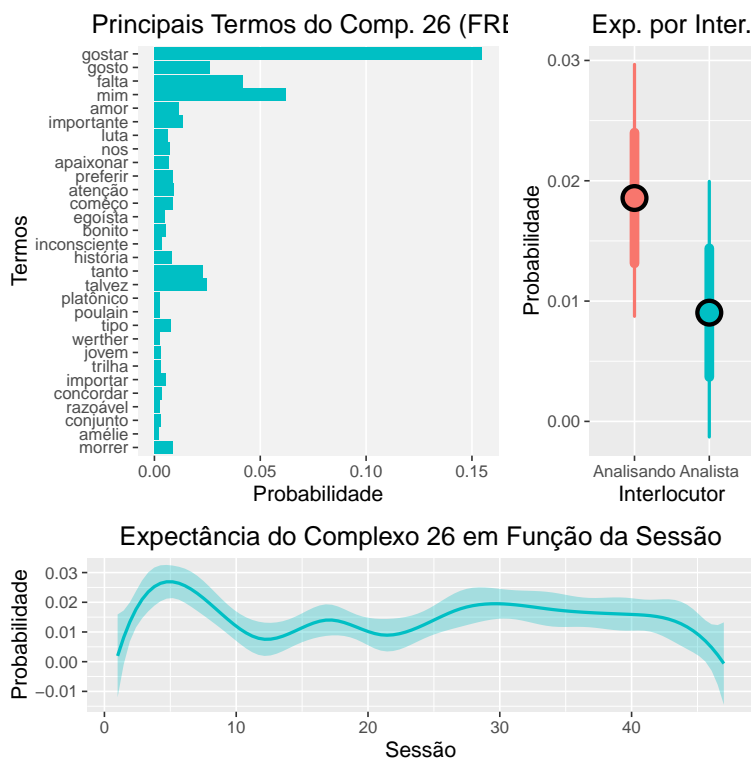


**Figura 5.8:** Gráfico das características do complexo 37.

(Sessão 3, Turno 32)

A apresentação por relevância no modelo inverte a ordem de aparecimento dos turnos de fala durante a sessão. A fala de Ivan remete a suas fantasias de morar sozinho, isolado, em algum lugar ermo. Ele confessa que seu maior temor nesta situação imaginária seria ficar sozinho com seus próprios pensamentos, porque eles insistem em retornar à Nina, atribuindo a ela toda e qualquer motivação para suas ações. O pensamento obsessivo sobre Nina é exemplificado em dois casos: no interesse por um concurso público e na motivação para a fantasia de abandonar o convívio com os outros. Na sequência, Ivan parafraseia Thoreau, comentando que, quando alguém se vê numa situação diferente daquela na qual se encontra, estará perdido entre duas situações, com o dobro de dificuldade para sair delas.

A intervenção do analista retoma o sentimento de confusão que foi mencionado logo na primeira entrevista, situando-o no contexto desta percepção de dupla localização, e, em particular, a dificuldade de sair dela. O exemplo dessa divisão é retomado da relação com Helena, pois Ivan confessa ter tido dificuldades nas primeiras relações sexuais com ela porque simplesmente não conseguia deixar de pensar em Nina. De fato, a relação de Ivan com as mulheres é marcada muito mais pelos seus pensamentos e devaneios a respeito de seus interesses amorosos do que pela convivência. O complexo 26, positivamente correlacionado com o complexo 37, trata da posição platônica que marca as paixões de Ivan. Utilizando o critério FREX para ordenar os termos, a figura Figura 5.9 elenca os termos *gostar*, *falta*, *amor*, *apaixonar* e *platônico*, além de referências a personagens românticos como Werther e Amélie Poulain.



**Figura 5.9:** Gráfico das características do complexo 26.

O complexo sobre a natureza platônica dos amores de Ivan aparece com maior probabilidade em seus próprios turnos de fala. Sua frequência ao longo do tratamento é relativamente menor que outros temas, mas permanece constante e com poucas alterações. Sua distribuição temporal difere daquela do complexo 37, que possui um pico nas primeiras dez sessões e volta a aparecer com força nas últimas sessões, para enfim voltar a diminuir. O turno de fala mais representativo do complexo 26 pertence à quinta sessão, no mesmo momento em que o complexo sobre as mulheres se encontrava bastante ativo.

Ivan: Não, não é bom ser incomodado. Era bom estar com ela [a Nina], as outras coisas. Ser incomodado, era, talvez, aceitável, por causa das outras coisas. Tinha uma compensação, de certa forma. Então talvez eu não me importasse tanto. [...] Eu pensei em outra coisa agora, que eu só tinha pensado em relação a minha avó, que morreu em 2010, e o meu avô, que morreu no ano passado. Um trequinho de *Os Sofrimentos do Jovem Werther*. Tem um trecho que ele fala que nada nos torna mais necessários aos outros do que o amor que lhes temos. E isso me chamou muito a atenção a primeira vez que eu li lá em 2007. [...] Eu tinha meio que presenciado a morte existencial dele [do avô], digamos assim. [...] Era ruim morar lá com ele. Mas ele sentia bastante por mim. Ele dizia que eu era o neto preferido, e falava um monte. Ficava falando bem de mim o tempo todo. Tanto que eu me sentia mal por causa disso, de um jeito diferente. E eu percebi isso de novo. Talvez com Nina tenha sido a mesma coisa. Ela gostava bastante de mim, e talvez seja isso que mais faça falta. Embora eu também gostasse dela. (Sessão 5, Turno 33)

A fala de Ivan ilustra com clareza o tom idílico dado por ele às relações amorosas. De saída, ele comenta sobre o mal-estar em se sentir incomodado por Nina, tema que emergiu principalmente no complexo sobre o sentimento de ira. Em seguida, associa sobre a morte do avô e como se sentia querido por ele, citando Goethe e, não por acaso, um excerto do *Os sofrimentos do Jovem Werther*, para falar da saudade que sente dos sentimentos ternos que lhe foram dirigidos, tanto pelo avô quanto por Nina. A alusão ao livro de Goethe também permite retomar as ideias suicidas de Ivan

e seu entendimento sobre o suicídio do avô, supostamente por solidão. Como é de conhecimento comum, à publicação de *Werther* é atribuída uma onda de suicídios, inspirados no desfecho trágico do triângulo amoroso que se desenvolve no romance. A citação, portanto, apesar de partir do amor idealizado pelas mulheres, serve de alusão às lembranças da morte do avô e a suas próprias ideias sobre a morte e o morrer, como capturado nos termos do complexo.

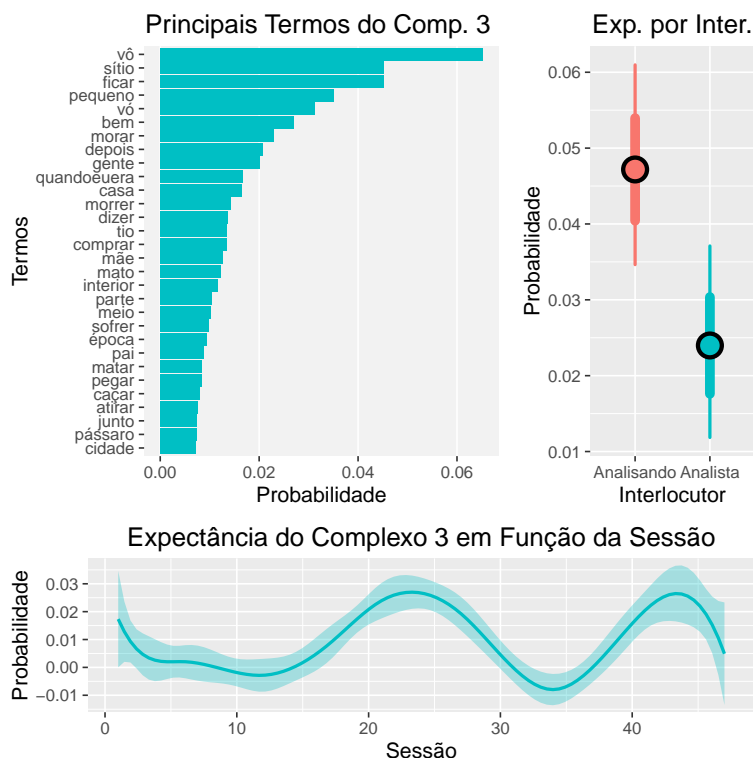
### 5.2.6 Memórias do avô

A importância da afetividade do avô vai além de sua presença marginal em um complexo. Ivan se refere constantemente às suas lembranças, ao tempo em que morou em seu sítio e à sua recente perda, a ponto de estas associações tomarem um dimensão própria e constituírem o complexo 3. Como pode ser verificado na Figura 5.10, os principais termos são *vô*, *sítio*, *pequeno*, *vó*, *morar*, *quandoeuera*, *casa* e *morrer*. As vivências com o avô no período que antecede a mudança para a capital, durante seus estudos de graduação, estão misturadas no complexo a outras lembranças mais primitivas, como expresso pelos termos “quando eu era pequeno...”, e por memórias recentes relativas à morte e ao sofrimento pelo qual Ivan passou por sua perda.

O modelo detecta dois picos na probabilidade de ocorrência deste complexo. O primeiro pico, que acontece mais ou menos na metade do tratamento, coincide com o aniversário do falecimento do avô. O segundo pico, por sua vez, coincide com o período do fim de ano e emerge nas associações sobre as saudades deixadas pelo fato de que as festas familiares para comemorar o Natal e o Ano Novo costumavam acontecer no sítio. Se compararmos a expectância temporal do complexo com os complexos anteriores, podemos verificar que as associações sobre o avô aumentaram durante um período em que a probabilidade de outros complexos diminuiu, referendando mais uma vez a importância desta temática num período em que as lembranças de Ivan eram prontamente recordadas em função dos eventos correntes.

É a esse período de lembranças pesarosas que pertence o turno de fala no qual o complexo 3 é o mais prevalente, de acordo com as estimativas do modelo:

Ivan: [...] Sim, eu fiquei assustado. Porque o passarinho estava na árvore e eu atirei nele, e daí ele caiu. Só que ele estava pulando e se mexendo. Eu fiquei meio



**Figura 5.10:** Gráfico das características do complexo 3.

assustado, porque ele não morreu. Ele estava sofrendo ali, e o vô disse, atira de novo, então. Quando ele caiu, eu fiquei alegre, eu fiquei feliz, porque eu tinha acertado o passarinho. Ele caiu e eu fiquei assustado porque ele estava vivo, sofrendo. E agora, o que que eu faço? E o vô disse, atira de novo! Porque o passarinho parar de sofrer, de certo. E daí eu atirei de novo, e então ele morreu, eu acho, ou atirei mais um vez, eu não me lembro. Mas foi meio assustador. Veio uma alegria no começo e depois uma certa tristeza. [...] Embora eu lembro de outras vezes que... [...] Quando o vô comprou o sítio, ele não caçou mais. E eu também... Não, eu cacei algumas vezes. Uma vez, eu acho. Porque um amigo de

Sevastopol quis ir e eu fui junto. Mas eu não estava mais para a caça. [...] Mas eu não estava nem aí para o passarinhos. Eu achava divertido ir para o meio do mato e ficar andando de um lado para o outro. [...] E a gente ia para o mato, também, e às vezes a gente chegava num lugar que tinha bastante grama, e tinha bastante árvore. Era bem fechado, só que dava para ver o sol entrando, também. A gente deitava no chão e ficava olhando para cima. E era legal, porque dava para ver os galho das árvores e o céu azul. (Sessão 23, Turno 15)

Neste excerto, Ivan se recorda de uma memória de infância, quando saiu para caçar com seu avô pela primeira vez. O júbilo por ter conseguido acertar o alvo na primeira tentativa, como uma demonstração para seu avô, é contrastado com o terror ao verificar que o pássaro, mesmo alvejado, continuava vivo. A justificativa do júbilo parece ser explicada na sequência de associações: ainda que não seja possível descartar o prazer sádico pela morte do pequeno animal, é dos passeios pelas regiões ermas do interior com o avô que Ivan retira a maior parte de sua satisfação destas atividades.

### 5.2.7 *Wanderlust*

Os passeios com o avô pelo interior de Sevastopol e a longa permanência no sítio antes da mudança para Moscou deixaram em Ivan uma marca indelével, que começou a se manifestar na adolescência e permanece um importante elemento de sua economia libidinal. Trata-se de um forte desejo de sair sem direção pelo mundo, atualizado em fantasias com diversas variações do tema. Em algumas dessas variações, o principal objeto em jogo é se livrar de uma situação presente a respeito da qual Ivan não consegue tomar uma posição decisiva – o restante da fantasia serve apenas de mecanismo para realizar esse desejo. Em outras variações, o foco maior é o sentimento de solidão e as viagens servem de instrumento para encontrar e conhecer outras pessoas, geralmente caracterizadas como mais compreensíveis. Algumas dessas pessoas são bons amigos, espíritos irmãos com quem Ivan consegue compartilhar suas ideias. Outras, melhor caracterizadas, são mulheres especiais que se tornam suas companheiras, pelo menos por um período de tempo. Nesses casos, em geral, Ivan assemelha seus devaneios ao mito do herói, construindo também a ideia de retorno ao lugar de onde saiu. Por fim, algumas edições da fantasia mudam seu objetivo final: em vez



de caminhar pelo mundo, Ivan se isola da sociedade em algum lugar ermo, geralmente imaginado a partir das memórias da infância no interior de Sevastopol.

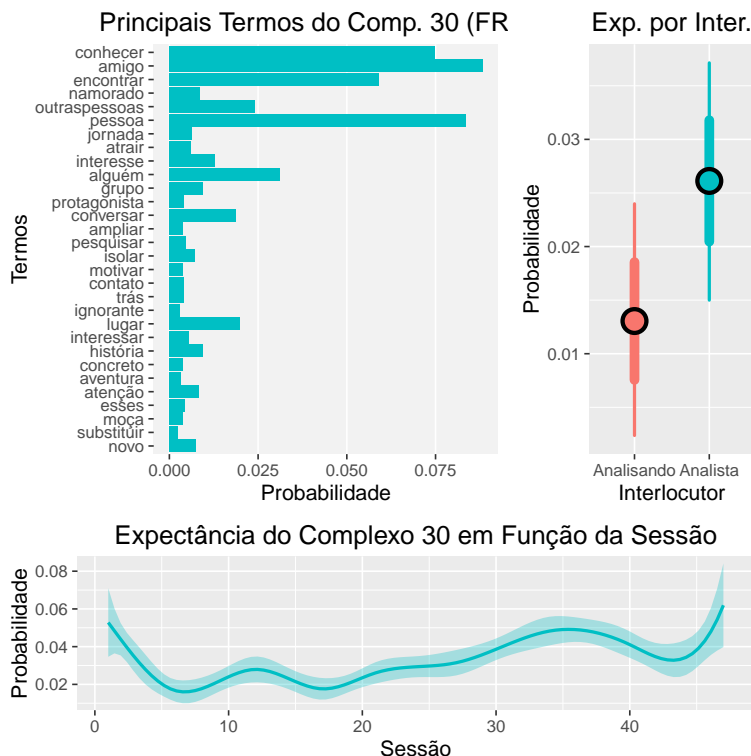
As características do complexo 30, que condensa esses elementos, são apresentadas na Figura 5.11 utilizando a estratégia FREX para determinar os principais termos. Pelas estimativas do modelo, o complexo é consideravelmente mais comum nos turnos de fala do analista, ainda que sua probabilidade nas associações do analisando não sejam baixas. Sua variação ao longo do tempo indica que houve uma rápida redução nas primeiras sessões, aumentando de maneira constante, mesmo que lenta, ao longo do tratamento. Os principais termos identificados são *conhecer*, *amigo*, *encontrar*, *namorado*, *outras pessoas*, *pessoa* e *jornada*. O conjunto de palavras captura bem os objetivos propostos nas fantasias, de conhecer outras pessoas, e a ideia de que isso aconteceria durante uma aventura.

Os seis primeiros turnos de fala mais representativos pertencem ao analista. Trazemos o primeiro, indicando o teor da intervenção que aborda o complexo 30, e complementamos com dois turnos de fala de Ivan: um deles contemporâneo à intervenção apresentada, e outro do início do tratamento, no primeiro momento em que este tópico aparece com grande probabilidade.

Analista: Parece que esse é o ponto das fantasias que trazias. A ideia de sair e encontrar outras pessoas sempre vinha caracterizado pela estrutura mitológica do herói, que retorna com os espólios da aventura. E tu vias nesse espólio nem tanto uma coisa material, ou trazer essas pessoas junto contigo fisicamente, mas se apropriando da experiência e habilidade dessas pessoas. [...] Em tua fantasia, não é tanto esse desejo de liderança mas se apropriar do que as outras pessoas têm de especial. (Sessão 36, Turno 36)

Ivan: Não sei, foi uma coisa que me marcou desde pequeno. Sempre me chamava a atenção aqueles filmes de aventura, onde o protagonista conhecia um monte de pessoas e cada um tinha uma característica diferente. (Sessão 36, Turno 5)

Ivan: Sabe, quando eu era adolescente, era um pouco diferente. Eu sempre voltava, quando saía. Eu saía e voltava. Não como se fosse um rito de passagem, mas algo maior. A jornada do herói. Que eu só vim a conhe-



**Figura 5.11:** Gráfico das características do complexo 30.

cer depois, lendo Campbell. Seria mais ou menos isso. Depois eu percebi essa coisa toda. Era como se eu sáísse em busca de algo, encontrasse e voltasse e tivesse algo novo. (Sessão 3, Turno 56)

A associação de Ivan logo na terceira sessão reconhece a grandiosidade mitológica de suas fantasias de aventuras e jornadas – razão, talvez, para nunca levá-las a cabo de fato, pelo menos não de maneira próxima ao que foi idealizado. A fala da sessão 36, por sua vez, vai na direção de buscar as experiências originárias das fantasias e do desejo de perambular. Não ocorre a Ivan, no momento, as aventuras que teve com o avô durante a infância, e sua justificativa permanece ainda na constatação que já se via atraído por esse tipo de narrativa quando criança.

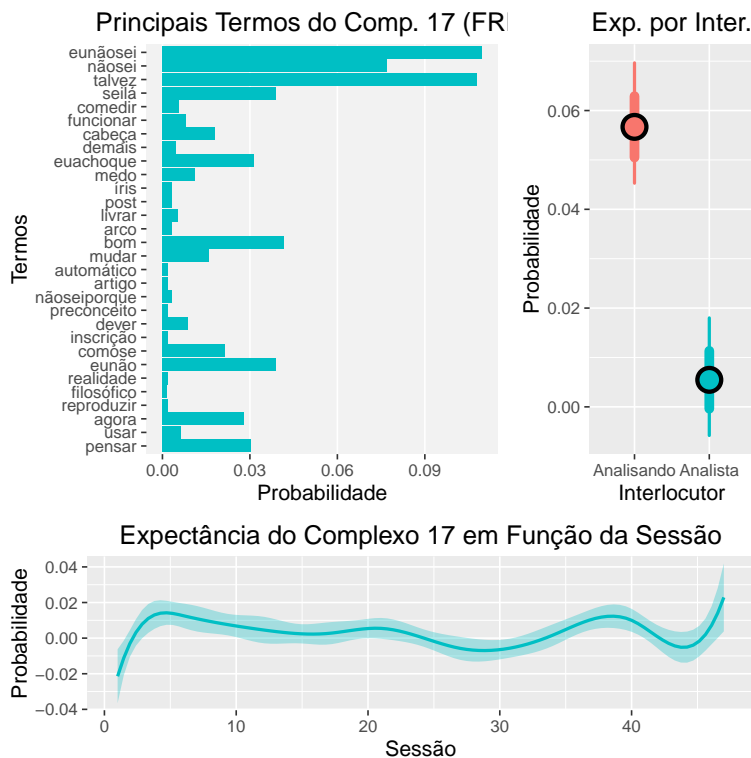
A interpretação do analista, por fim, retoma, em um momento mais avançado do tratamento, a intuição fornecida pelo analisando durante as primeiras entrevistas. Mas a referência ao retorno do herói é tornada mais precisa em termos das transformações que Ivan fantasia passar em sua aventura legendária. Em particular, a interpretação remete à introjeção de características desejáveis dessas pessoas especiais que Ivan encontraria em sua caminhada. Se juntarmos à interpretação a fala representativa do complexo 26, que remete à citação do jovem Werther, há indícios de que o objeto central destas fantasias é o desejo de se perceber amado pelos outros.

### 5.2.8 Barreiras à associação livre

Os complexos abordados até agora têm como característica refletirem núcleos temáticos abordados por Ivan durante seu tratamento, indicando por meio da magnitude de probabilidade quais termos estão mais semanticamente próximos e podem ser substituídos uns pelos outros ou coocorrerem num mesmo contexto de enunciação. Muitos outros complexos identificados pelo modelo – apresentamos apenas 12 dos 50 utilizados para ajustar o modelo ao *corpus* tratado – não são prontamente interpretáveis, pois juntam termos cujos padrões de coocorrência são fortuitos ou representam organizações não imediatamente óbvias para o entendimento humano.

Mas há outros complexos que não parecem representar núcleos de significação. Esses complexos estão mais relacionados a determinadas posturas enunciativas, caracterizadas pela utilização de termos estereotipados e com pouco valor semântico. Em particular, o complexo 17 parece indicar a atitude usualmente denominada de *resistência* durante o processo de análise, em seu sentido original do momento em que as associações faltam ao analisando e ele se mostra incapaz de seguir falando (Freud, 1984f), ou, ainda, como o que interrompe o desenvolvimento do processo psicanalítico (Freud, 2001). Independente da definição escolhida, o fato é que o modelo consegue capturar a utilização reiterada de expressões que demarcam a interrupção do processo de livre associação, principalmente sob a forma de *dúvida* ou *incapacidade*. Como pode ser observado na Figura 5.12, os principais termos que indicam essas táticas são *eunãosei*, *talvez*, *nãosei*, *bom*, *seilá*, *eunão* e *euachoque* – o que sugere uma posição de resistência pouco reativa, interpondo entraves à recordação, principalmente.

Como esperado, o complexo que demarca a resistência apa-



**Figura 5.12:** Gráfico das características do complexo 17.

rece com probabilidade cerca de seis vezes maior nos turnos de fala do analisando – em verdade, nos turnos do analista, a probabilidade de aparecimento deste complexo é quase zero. Também é interessante notar que há pouca variação temporal em sua ocorrência, sugerindo que, afora um curto intervalo nas cinco primeiras sessões, a resistência se manteve mais ou menos constante ao longo de todo o tratamento. Dois excertos ilustram com precisão esse momento de hesitação no processo associativo:

Ivan: Sim. [...] Eu acho que sim. Eu acho que só estou sofrendo por antecipação. Talvez nem por antecipação, porque eu não vou sofrer depois. Não sei, eu acho que é só isso. (Sessão 13, Turno 39)

Ivan: Sim. Eu acho que sim. Tem um monte de coisa que

veio na minha cabeça agora. [...] Eu não sei por onde eu começo. Essa coisa da terapia, que eu não entendo bem como é que funciona. Eu não sei se você poderia me explicar? Como é que funciona? Digo, tem várias linhas, com várias teorias diferentes, e o resultado é para ser mais ou menos o mesmo? (Sessão 23, Turno 59)

Os dois turnos de fala, ocorridos com dez sessões de diferença, mostram duas táticas diferentes de manifestação da resistência. A primeira fala é um bom exemplo de como a dúvida pode ser utilizada para interromper o fluxo associativo, enfraquecendo uma afirmação inicial até esvaziá-la de qualquer valor. O segundo, por sua vez, inicia com a dúvida para, em seguida, impedir a associação sob a justificativa de estar sobrecarregado de pensamentos. A natureza de resistência dessa justificativa se mostra clara quando, em vez de escolher um ponto para iniciar as associações, prefere fazer o analista explicar detalhes sobre o funcionamento do tratamento.

Já que o complexo 17 trata justamente de momentos em que a associação livre falha, podemos utilizar os parâmetros do modelo para identificar junto de quais outros complexos ele tenta a ocorrer. Em outras palavras, podemos identificar quais complexos estão mais diretamente relacionados com o fenômeno da resistência nas associações de Ivan. Os complexos mais fortemente correlacionados são, dentre os que abordamos até agora: o complexo 31, que trata do retorno à Sevastopol; o complexo 18, que reúne termos relacionados à estratégia de evitação de conflitos interpessoais; e o complexo 26, que aborda as ideias idílicas de Ivan sobre o amor. O complexo que demarca a resistência também coocorre com o complexo 42, que trata de uma série de termos relacionados a atividades cotidianas e que podem ser considerados como assuntos triviais – uma diferente estratégia de resistência, possivelmente.

## 5.3 Outras inferências do modelo

Outros complexos identificados pelo modelo poderiam ser abordados, mas sua relevância para a apresentação do caso de Ivan é questionável. Por exemplo, o complexo 42, recém-mencionado, aborda as atividades e fazeres cotidianos, como os hábitos de dormir, o que faz assim que acorda, tarefas que realiza ao longo do dia. Alguns complexos reúnem informações sobre os amigos que Ivan encontra na universidade e com quem divide o lugar onde mora, em conjunto

com as atividades que costumam realizar juntos, como almoçar ou limpar os cômodos da casa (complexos 36 e 22). Por sua natureza de fala vazia, pouco acrescentariam a uma melhor caracterização do caso<sup>5</sup>.

No lugar de apresentar o restante dos complexos em detalhe, encerramos o capítulo com duas informações adicionais que podem ser extraídas do modelo ajustado: a probabilidade esperada de cada complexo em todo o *corpus* e um grafo de proximidade entre os diferentes complexos, com base na matriz de covariância  $\Sigma$ . Desta maneira, além de retomar os complexos já apresentados dentro do contexto mais geral das inferências do modelo, apresentamos rapidamente alguns aspectos dos complexos que não foram descritos em detalhes.

### 5.3.1 Probabilidade dos complexos

Utilizando a matriz  $\Theta$ , que reúne a proporção esperada de cada complexo em cada turno de fala, podemos computar o valor esperado marginal da proporção de cada tópico utilizando a média aritmética da proporção dos complexos específicos ao longo de todo o *corpus*. Essa estratégia, apesar de ser uma aproximação grosseira do valor esperado efetivo, é justificada como integração numérica, tendo em vista que o número de documentos é relativamente grande.

Utilizando o algoritmo FREX para criar rótulos significativos, a probabilidade dos complexos pode ser conferido na Figura 5.13, que os organiza em ordem decrescente de acordo com a magnitude da probabilidade esperada.

O complexo com maior probabilidade esperada versa justamente sobre as mostras de resistência de Ivan, conforme argumentamos na seção anterior. Os três complexos na sequência abordam temas que já discutimos anteriormente: a relação de Ivan com as mulheres, a questão do retorno à terra natal, e o sentimento de raiva que o incomoda. Os complexos seguintes abordam principalmente turnos de fala do analista, e não foram apresentados aqui em função de abarcaram diversos termos diferentes, de difícil interpretação sintética. Na sequência, novamente, reencontramos vários dos complexos que já abordamos acima.

Apesar da presença intrusiva de complexo que deixamos propositalmente de fora da exposição do caso, o valor esperado de cada

---

<sup>5</sup>Os demais complexos podem ser conferidos no *link* disponibilizado no Apêndice A sob a forma de uma visualização interativa.

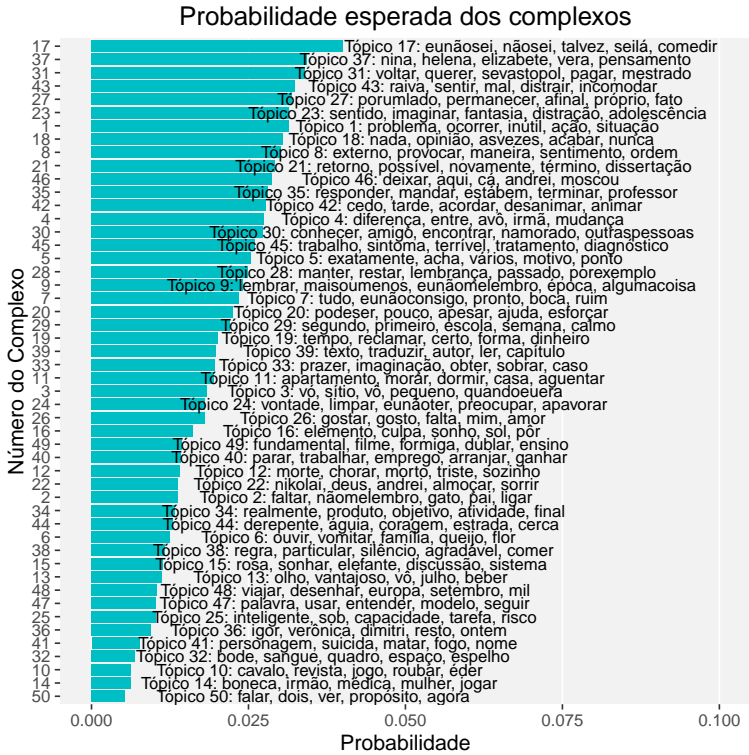


Figura 5.13: Probabilidade dos complexos

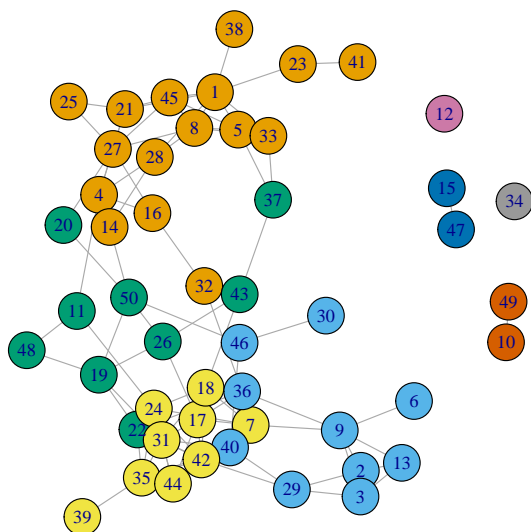
tópico dá uma ideia coerente da relevância de cada complexo no *corpus*. A coerência é justificada pela intuição que seguimos para a apresentação do caso, de que os principais temas estariam de alguma forma já presentes na entrevista inicial. Nesse sentido, as inferências do modelo mais uma vez são corroboradas por algumas expectativas externas, como nossa expectativa intuitiva dos principais núcleos temáticos.

5.3.2 Rede de proximidade entre complexos

Além de apresentarmos os principais termos de cada complexo, utilizamos a estrutura de correlação pressuposta pelo modelo utilizado para identificar possíveis complexos próximos. Esta estratégia per-

mitiu tanto identificar complexos distintos, mas de conteúdo semântico próximo, quanto complexos que tendem a se suceder durante o processo de associação livre de Ivan.

Uma maneira de resumir a informação de correlação entre os complexos é por meio de um grafo. Cada vértice representa um dos complexos identificados pelo modelo, e as arestas ligam complexos cuja correlação positiva é comparativamente alta. A correlação é computada a partir da matriz de covariância  $\Sigma$  da distribuição Logit-Normal, utilizada para modelar o parâmetro  $\theta$  de proporção de cada complexo nos turnos de fala. A representação gráfica das correlações pode ser conferida na Figura 5.13. As diferentes cores indicam possíveis comunidades de complexos no grafo, de acordo com um algoritmo ganancioso para descoberta de comunidades.



**Figura 5.14:** Grafo de correlações entre complexos



É evidente a presença de complexos completamente isolados ou então ligado a apenas um outro. À comunidade em amarelo pertence o complexo de resistência (17) e de conversa fiada (42), acompanhados de diversos outros complexos importantes, como o retorno a Sevastopol (31), a estratégia de evitação (18), o término e tema da dissertação (35 e 39). Além de serem temas cruciais na análise de Ivan, eles também compartilham a característica de terem maior probabilidade nos turnos de fala do analisando.

A comunidade azul clara reúne algumas memórias infantis com a família (2), pessoais (6), e, em especial, o complexo sobre o avô (13). Também estão presentes memórias da escola (29) e do primeiro emprego (40), e a dificuldade em lembrar as recordações mais primeiras (9). Das questões mais contemporâneas, foram incluídos os complexos sobre as relações com os colegas de quarto (36), as fantasias de viagem e conhecer pessoas (30) e abandonar seus amigos na capital (46). Essa comunidade parece estar ligada às recordações de Ivan, principalmente, e às relações de amizades atuais, em menor escala.

A comunidade em cor verde, por sua vez, reúne complexos a respeito dos amigos da universidade (22 e 19), além de atividades realizadas com eles, como viagens (48). O tema da mãe e do lar também está incluído neste agrupamento (11), em conjunto com o complexo sobre a relação com as mulheres (37) e o sentimento de raiva (43). O principal teor, portanto, parece ser a relação conflituosa com a mãe e as mulheres, expressas ambigualmente no sentimento terno ou apaixonado, mas também manifesto sob a forma de ira.

Os vértices da comunidade laranja-claro, por fim, remetem a complexos ligados principalmente aos turnos de fala do analista – a maioria deles de difícil interpretação, como já mencionado, com a possível exceção do complexo 21, que aborda o término da dissertação em conjunto com o retorno a Sevastopol.

## 5.4 Considerações

A apresentação do caso de Ivan Ivanovitch está longe de poder ser tomado como um *roman à clef*. Diferentemente das grandes psicanálises clássicas publicadas por Freud, não expusemos a história de vida do analisando e seus achados numa narrativa coesa e coerente. Pelo contrário: a ordem da apresentação seguiu de perto as inferências permitidas pelo modelo de tópico que aplicamos às transcrições

das associações livres e intervenções do analista. Obviamente, para expressar melhor a maneira como os complexos capturam a proximidade, coocorrência e substituíbilidade dos termos que os compõem, lançamos mão de excertos de falas e de algum contexto do que estava acontecendo no tratamento.

Mas não nos deixemos enganar pela miríade de caminhos que a interpretação dos complexos e de outros parâmetros do modelo poderiam ter à luz da teoria psicanalítica. O que de fato realizamos foi apresentar, de maneira densa e descritiva, os resultados numéricos obtidos a partir do modelo ajustado aos dados. Quaisquer interpretações extras que possamos ter tecido ao longo da apresentação do caso precisam ser desculpadas: é o caso em que a explicação de um resultado pode levar ao acréscimo de informações que não estavam contidas nele. A opção seria decidir por uma apresentação mais dura, puramente numérica, mas esteticamente pouco atraente.

Com isso, esperamos que o caso de Ivan possa ser tomado como uma *prova de conceito* (*proof of concept*) de que alguns aspectos da Psicanálise podem ser formulados formalmente como modelo de tópicos. Os resultados obtidos são profícuos porque o formalismo, longe de ser somente uma miragem de rigor, permite expressar as ideias de uma maneira mais clara e sistemática – certamente não é nossa intenção aumentar ainda mais o hermetismo que envolve diversas formulações teóricas da Psicanálise no contexto atual. Além disso, como mostramos no capítulo anterior, o modelo também permite fazer predições e testar hipóteses de maneira rigorosa. Por fim, os resultados também são úteis, pois, ao facilitar a comunicação de elementos relevantes da experiência clínica levando em consideração a totalidade de informações disponíveis e com menos viés de seleção do pesquisador, permitem estabelecer um elo mais claro e objetivo entre a experiência analítica e a maneira como ela pode informar ou revisar a teoria da Psicanálise.

## Conclusão

Nosso percurso iniciou com o questionamento do lugar da experiência clínica na construção, avaliação e revisão da teoria psicanalítica. Alinhamos nossa posição com os críticos que exigem da Psicanálise um suporte empírico mais confiável, e renegamos a possibilidade de construção de conhecimento a partir do uso de excertos deslocados da experiência clínica, invocados sem justificativa para referendar a verdade da teoria que era conhecida desde o início. Nossa crítica, todavia, seguiu por um caminho relativamente diferente: ao invés de buscar inovações metodológicas que oferecessem o fundamento desejado, focamo-nos na distinção entre modelos analógicos e digitais, identificando a primeira categoria como dominante absoluto na teoria psicanalítica. Nesse ponto, apresentamos uma nova exigência: se não trabalharmos no sentido de construir um mínimo de formalização de alguns aspectos da teoria psicanalítica, nenhum evidência empírica poderá ser avaliada de maneira sistemática.

E foi a isso que nos propomos nos primeiros capítulos. Elegermos as associações livres como o objeto de interesse mais imediato para a formalização, justamente por ser a pedra angular da prática clínica e por constituir, portanto, toda base de evidência que os psicanalistas utilizam para a construção de suas teorias. Evitaríamos, com isso, o risco de construir uma formalização vazia, que apenas colocasse ordem na casa dos conceitos sem apresentar nenhuma aplicabilidade para justificar a ponte entre a clínica e a teoria. Partimos de uma revisão dos modelos analógicos das associações livres, para, em seguida, descrevermos em detalhes os modelos de tópicos e sua aplicação para a análise de dados textuais. Conjugando as contribuições freudianas com elementos da teoria da probabilidade, construímos um objeto modelo para as associações livres e sua respectiva teoria específica sob a forma de um modelo de tópicos.

Para demonstrar a utilidade do modelo, conduzimos um estudo de caso especialmente voltado para esse fim. As 47 sessões da psicanálise de Ivan Ivanovitch foram integralmente registradas em áudio e transcritas para a composições do *corpus* utilizado no ajuste e análises do modelo. Antes de efetuar qualquer inferência, contudo,

buscamos estabelecer critérios para sua validade externa. O ajuste escolhido foi avaliado em uma tarefa de predição, utilizando os parâmetros ajustados para predizer variáveis de critério, e comparando suas predições teóricas ao resultado efetivo de um conjunto de testes de associação de palavras. Com relação à tarefa de predição, os modelos de *random forest* foram capazes de predizer, com precisão, o autor de cada turno de fala e o momento do tratamento em que ele foi proferido, utilizando como preditor apenas o vetor de proporção de tópicos por documento. O resultado obtido no teste de associação de palavras, por sua vez, contrastado o modelo proposto a um modelo mais simples, sem pressupor a estrutura do modelo de tópicos, não foi tão decisiva quanto gostaríamos. Mesmo assim, a capacidade preditiva do modelo foi referendada, pelo menos em relação ao modelo de referência. Tendo estabelecido alguma confiança nas inferências, apresentamos os complexos descobertos pelo modelo nas associações de Ivan e nas intervenções do analista, avaliando sua dinâmica temporal e sua prevalência na fala de cada interlocutor. O modelo também nos permitiu apresentar informações qualitativas de uma maneira fundamentada: os turnos de falas apresentados, uma amostra real dos documentos componentes do *corpus*, foram identificados a partir dos parâmetros do modelo.

Nossa tese, portanto, de que *as associações livres podem ser representadas formalmente como modelos de tópicos* foi satisfatoriamente corroborada no contexto das inferências realizadas para o estudo de caso. As avaliações do ajuste e o resultado obtido para as hipóteses preditas, porém, indicam que há ainda um longo caminho para melhor adequação do modelo. Discutimos, agora, algumas das limitações do estudo e apontamos, em seguida, suas possíveis aplicabilidades em estudos futuros.

## Limitações

Uma primeira objeção que pode ser feita ao estudo é o fato de ele avaliar o modelo exclusivamente a partir de um estudo de caso. Esta é uma limitação óbvia, porém importante, e que merece algumas considerações antes de ser considerada em maior profundidade. Primeiramente, não é o caso que a amostra do estudo seja igual a um. Lembremos que as inferências feitas pelo modelo dizem respeito aos complexos constitutivos do psiquismo de um sujeito em particular, tal qual são representados no modelo. Nesse sentido, não foi um objetivo do estudo promulgar a generalização de quaisquer resultados

para uma população. Pelo contrário, a ideia de utilizar o estudo de caso foi uma prova de conceito que se mostrou frutífera. O modelo permitiu fazer predições e inferências que foram razoavelmente verificadas, baseadas em um volume consideravelmente grande de dados: 3.172 turnos de falas, 3.007 termos e 82.457 palavras na versão tratada do *corpus*, além dos 91 pares de palavras do teste de associação. Nesse sentido, os dados são suficientes para o escopo proposto.

Mesmo assim, é importante que seja verificada a replicabilidade dos resultados com outros estudos de caso. Nada impede, também, que o modelo seja utilizado em estudos de casos múltiplos, o que permitiria avaliar sua validade considerando a diversidade de sujeitos participantes. Tendo estabelecido um protocolo consideravelmente bem descrito neste estudo, a maior dificuldade agora é conseguir mais transcrições de sessões clínicas, já que a prática de gravação em áudio e transcrição não é comum, especialmente entre psicanalistas.

O custo dos dados necessários para a utilização do modelo é outra limitação importante do estudo. Diferente dos instrumentos usuais utilizados para avaliação de desfechos em contextos psicoterapêuticos, como questionários e inventários psicométricos autoaplicados, a transcrição de entrevistas consome muito tempo e ainda não pode ser feito de maneira automatizada. Obviamente, uma equipe grande de pesquisadores com recursos abundantes poderia contratar profissionais especializados para acelerar a obtenção dos dados, mas isso não resolve o problema do tempo necessário para obtenção de *corpora* grandes o suficientes para aplicar o algoritmo do modelo.

Além disso, como vimos no tratamento dos dados, somos obrigados a nos desfazer de um volume consideravelmente grande das transcrições, cuja obtenção foi custosa. A saída para não desperdiçar esses dados é complexificar a representação das informações sintáticas no modelo, como no modelo de tópicos que integra informações sintáticas e semânticas de Griffiths et al. (2004). O maior problema de um modelo com representações mais complexas é a dificuldade de sua implementação e a maior exigência de poder de processamento computacional. Nosso modelo, implementado sob a forma do STM, torna as computações mais simples assumindo a independência dos termos e não diferenciando termos de uso sintático e termos com conteúdo semântico. Mesmo assim, o desempenho na descoberta de núcleos temáticos nem sempre é necessariamente melhor em um modelo mais complexo, já que ele precisa ajustar mais

parâmetros, o que exige mais dados para estimações precisas.

Além das limitações e simplificações que se fizeram necessárias na implementação do modelo e do seu algoritmo, também não é plenamente satisfatório utilizar a distribuição marginal dos termos como modelo comparativo da performance preditiva do modelo de tópicos. Nesse caso, porém, estamos de mãos atadas: na ausência de outros modelos digitais para caracterizar as associações livres ou algum outro modelo mais genérico da interação em psicoterapias, somos obrigados a recorrer a modelos nulos com pouco ou nenhum valor prático – ainda que muitas técnicas de análise de conteúdo da transcrição de psicoterapias se baseie justamente na frequência de termos de um dicionário definido *a priori*.

A análise preditiva posterior sugere que o volume de dados utilizados para ajustar o modelo ainda não está num nível ótimo. Certamente, já foi suficiente para a descoberta de estruturas importantes nas associações, mas alguns dos tópicos ainda parecem ser fruto do superajuste do modelo a um conjunto pouco informativo de dados. Este é um problema, entretanto, que assola toda forma de pesquisa empírica de processamento de linguagem natural por computador. Como bem ilustrado pela lei de Zipf, tendemos a fazer uso de um vocabulário variado em que a maioria das palavras são utilizadas com baixíssima frequência, enquanto outros termos, pouco significativos, aparecem em abundância. A única solução possível, nesse caso, é aceitar a maldição da dimensionalidade e obter mais dados para conseguir melhores ajustes.

Por fim, não podemos perder de vista que, neste trabalho, a proposição do modelo, construção das hipóteses, condução do tratamento e análises dos dados foram todos conduzidos pela mesma pessoa. Sem elaboração de qualquer tipo de “cego” para a pesquisa, o risco dos resultados serem frutos de viés de confirmação é considerável. Em nossa defesa, podemos apenas aceitar a possibilidade da ocorrência deste viés e indicar que procuramos ser o mais claro e transparente possíveis na condução do tratamento de dados, avaliação das hipóteses e apresentação do caso.

## Aplicações e perspectivas

A utilização de modelos de tópicos, de maneira geral, pode permitir a diversas áreas da Psicologia trabalhar de maneira mais rigorosa com dados de natureza mais flexível e pouco estruturada, como dados textuais. A pretensão proposta por Blei (2013), inclusive, é que

pesquisadores de áreas que lidam com dados textuais possam construir modelos específicos, ao invés de ficar reutilizando sempre os mesmos modelos genéricos, codificando nesses modelos informações importantes previstas por suas teorias. Foi neste espírito que propusemos nosso modelo de tópicos, a despeito da dificuldade em implementá-lo exatamente como gostaríamos. Mesmo assim, os resultados obtidos neste estudo servem para ilustrar a potencialidade do uso de modelos de tópicos para a Psicologia de modo geral, mesmo que só como uma ferramenta heurística para tratamento de dados.

No nosso caso, em particular, acreditamos ter demonstrado que é possível construir modelos formais para as teorias analógicas que estudamos, e utilizar esses modelos para realizar inferência baseadas em evidências mais fortes do que a seleção enviesada de excertos de dados. Dessa formalização inicial, como já notamos, podemos elaborar modelos mais complexos: acrescentar informações mais práticas, como distinção entre termos sintáticos e semânticos, ou avançar no sentido da estruturação de outros elementos da teoria, como, por exemplo, o acréscimo de elementos distintos que permitam representar o processo primário de maneira explícita. Para isso, contudo, precisamos construir com clareza representações formais desses processos. O ponto principal é que a modelagem probabilística é uma ferramenta flexível e rigorosa que permite representar formalmente vários aspectos de teorias psicológicas, sem perder de vista sua conexão com os dados empíricos.

Uma outra possibilidade de pesquisa que pode fazer uso já do modelo aqui sugerido é a identificação de estruturas associativas comuns entre pacientes com diferentes tipos de psicopatologia. Esperamos que a maioria dos complexos diga respeito à singularidade de cada sujeito, mas o modelo permite também identificar complexos associativos utilizados por vários casos diferentes. Esse tipo de investigação poderia permitir avaliar empiricamente outros elementos teóricos da Psicanálise, em especial aqueles ligados à descrição da histeria e da neurose obsessiva. Mesmo fora de uma perspectiva psicanalítica, o conteúdo de tópicos e sua dinâmica temporal poderiam ser avaliados no tratamento de diferentes tipos de psicopatologias, permitindo inclusive a criação de marcadores diagnósticos baseados nos termos utilizados durante as sessões.

Por fim, os parâmetros inferidos pelo modelo podem ser definidos como desfecho para avaliação de processo do tratamento, inclusive em casos de estudos experimentais, de maneira análoga ao que Roberts, Stewart, e Airolti (2016) propõem para as ciências

sociais. Entrevistas avaliativas podem ser analisadas de maneira automática, dispensando a necessidade de juízes externos para codificar as respostas. Esta perspectiva é particularmente pertinente à avaliação de processo em Psicanálise, pois uma das críticas de psicanalistas às avaliações mais rigorosas é que as medidas utilizadas são artificiais e podem não corresponder ao que teoricamente se espera na condução de uma análise. Utilizando o modelo de tópicos, outras medidas de desfecho, teoricamente informadas, podem ser sugeridas e utilizadas para mensuração de impacto.

## Momento de concluir

Os resultados razoavelmente satisfatórios das predições do modelo podem não ser suficientes para convencer pesquisadores em Psicanálise sobre a relevância da formalização para a condução de estudos mais rigorosos e conclusivos. Resta-nos a esperança de que, mesmo que os procedimentos e resultados possam ser alvo de críticas justificadas, sua proposição sirva pelo menos como deflagradora de uma discussão necessária sobre o valor da evidência que a clínica oferece para a sustentação da teoria psicanalítica.

Ampliando o escopo para a pesquisa em Psicologia, de modo mais geral, esperamos que a apresentação de um modelo formal no contexto da Psicanálise, vista usualmente como avessa ao uso de métodos quantitativos, sirva de inspiração para a criação de modelos teóricos que estabeleçam interfaces coerentes entre fundamentos teóricos e investigação empírica. Esse movimento é crucial para que as pesquisas em Psicologia abandonem a perspectiva criticada por Bunge (1974) – e ainda amplamente em voga – na qual o uso da matemática fica reduzido à aplicação de métodos estatísticos superficiais como forma de esconder a pobreza conceitual das teorias.



## Referências

- Aguiar, F. (2006). Questões epistemológicas e metodológicas em psicanálise. *Jornal de Psicanálise*, 39(70), 105–131.
- Aguiar, F. (2016, Março). Psicanálise e psicoterapia: o fator da sugestão no “tratamento psíquico”. *Psicologia: Ciência e Profissão*, 36(1), 116–129. Recuperado 2016-06-10, de [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S1414-98932016000100116&lng=pt&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1414-98932016000100116&lng=pt&nrm=iso&tlng=en) doi: 10.1590/1982-3703004102015
- Andrews, M., & Vigliocco, G. (2010, Janeiro). The Hidden Markov Topic Model: A Probabilistic Model of Semantic Representation. *Topics in Cognitive Science*, 2(1), 101–113. Recuperado 2014-06-24, de <http://onlinelibrary.wiley.com/doi/10.1111/j.1756-8765.2009.01074.x/abstract> doi: 10.1111/j.1756-8765.2009.01074.x
- Atkins, D. C., Rubin, T. N., Steyvers, M., Doeden, M. A., Baucom, B. R., & Christensen, A. (2012). Topic models: A novel method for modeling couple and family text data. *Journal of Family Psychology*, 26(5), 816–827. Recuperado 2014-06-24, de <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0029607> doi: 10.1037/a0029607
- Atkins, D. C., Steyvers, M., Imel, Z. E., & Smyth, P. (2014, Abril). Scaling up the evaluation of psychotherapy: evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science*, 9(1), 49. Recuperado 2014-06-24, de <http://www.implementationscience.com/content/9/1/49/abstract> doi: 10.1186/1748-5908-9-49
- Benoit, K., & Nulty, P. (2016). *quanteda: Quantitative Analysis of Textual Data*. Recuperado de <https://CRAN.R-project.org/package=quanteda> (R package version 0.9.4)
- Bishop, C. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Blatt, S. J., Corveleyn, J., & Luyten, P. (2006, Junho). Minding the Gap Between Positivism and Hermeneutics in Psychoanalytic Research. *Journal of the American Psychoanaly-*

- tic Association*, 54(2), 571–610. Recuperado 2015-02-17, de <http://apa.sagepub.com/content/54/2/571> doi: 10.1177/00030651060540021301
- Blei, D. M. (2012, Abril). Probabilistic Topic Models. *Commun. ACM*, 55(4), 77–84. Recuperado 2014-06-12, de <http://doi.acm.org/10.1145/2133806.2133826> doi: 10.1145/2133806.2133826
- Blei, D. M. (2013). Topic Modeling and Digital Humanities. *Journal of Digital Humanities*, 2(1). Recuperado 2014-06-28, de <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/>
- Blei, D. M., Griffiths, T. L., & Jordan, M. I. (2010, Janeiro). The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM*, 57(2), 1–30. Recuperado 2014-06-24, de <http://portal.acm.org/citation.cfm?doid=1667053.1667056> doi: 10.1145/1667053.1667056
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2016, Janeiro). Variational Inference: A Review for Statisticians. *arXiv:1601.00670 [cs, stat]*. Recuperado 2016-08-12, de <http://arxiv.org/abs/1601.00670> (arXiv: 1601.00670)
- Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning* (pp. 113–120). ACM. Recuperado 2015-03-14, de <http://dl.acm.org/citation.cfm?id=1143859>
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 17–35. Recuperado 2015-01-14, de <http://www.jstor.org/stable/4537420>
- Blei, D. M., & Lafferty, J. D. (2009). Topic Models. In A. N. Srivastava & S. Mehran (Eds.), *Text Mining: Classification, Clustering, and Applications*. Nova Iorque, Boca Raton, Londres: Chapman & Hall/CRC.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022. Recuperado 2014-06-24, de <http://dl.acm.org/citation.cfm?id=944937>
- Breuer, J., & Freud, S. (1984). Estudos sobre a histeria. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 2). Rio de Janeiro: Imago.
- Bunge, M. (1973). *Method, model and matter*. Boston: Reidel Publishing Company.
- Bunge, M. (1974). *Teoria e realidade*. São Paulo: Perspectiva.
- Dahl, H., Kächele, H., & Thomä, H. (1988). *Psychoanalytic process*

- research strategies*. Berlin, Nova Iorque, Londres, Paris, Tóquio: Springer-Verlag.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407.
- Doeden, M. A. (2010). *Examining the Utility of Topic Models for Linguistic Analysis of Couple Therapy* (Ph.D., Ann Arbor, United States). Recuperado 2014-06-24, de <http://search.proquest.com/docview/1322974463/abstract?accountid=26642>
- Edelson, M. (1985). *Hypothesis and Evidence in Psychoanalysis*. University of Chicago Press.
- Edelson, M. (1990). *Psychoanalysis: a theory in crisis*. Chicago: University of Chicago Press.
- Feinerer, I., & Hornik, K. (2015). *tm: Text Mining Package*. Recuperado de <https://CRAN.R-project.org/package=tm>
- Fienberg, S. E. (2006, Março). When did Bayesian inference become "Bayesian"? *Bayesian Analysis*, 1(1), 1–40. Recuperado 2015-03-11, de <http://projecteuclid.org/euclid.ba/1340371071> doi: 10.1214/06-BA101
- Fontes, F. F. (2015). A crítica à psicanálise: um capítulo censurado? *Psicologia em Revista*, 20(3), 446–459. Recuperado 2016-04-19, de <http://periodicos.pucminas.br/index.php/psicologiaemrevista/article/view/P.1678-9523.2014V20N3P446>
- Freud, S. (1979). Cinco lições de psicanálise. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (2ª ed., Vol. 11). Rio de Janeiro: Imago.
- Freud, S. (1982). Notas sobre um caso de neurose obsessiva. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (2ª ed., Vol. 10). Rio de Janeiro: Imago.
- Freud, S. (1984a). As transformações do instinto exemplificadas no erotismo anal. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (2ª ed., Vol. 17). Rio de Janeiro: Imago.
- Freud, S. (1984b). Conferência XXXV: A questão de uma Weltanschauung. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 22). Rio de Janeiro: Imago.
- Freud, S. (1984c). Construções em análise. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 23). Rio de Janeiro: Imago.

- Freud, S. (1984d). A dinâmica da transferência. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 12). Rio de Janeiro: Imago.
- Freud, S. (1984e). O inconsciente. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 14). Rio de Janeiro: Imago.
- Freud, S. (1984f). A psicanálise e a determinação dos fatos nos processos jurídicos. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 9). Rio de Janeiro: Imago.
- Freud, S. (1984g). Recomendações aos médicos que exercem a psicanálise. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 12). Rio de Janeiro: Imago.
- Freud, S. (1984h). Sobre o início do tratamento. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 12). Rio de Janeiro: Imago.
- Freud, S. (1984i). Uma breve descrição da psicanálise. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 19). Rio de Janeiro: Imago.
- Freud, S. (1984j). Uma nota sobre a pré-história da técnica de análise. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 18). Rio de Janeiro: Imago.
- Freud, S. (1984k). Um estudo autobiográfico. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 20). Rio de Janeiro: Imago.
- Freud, S. (1986). *Cartas a Wilhelm Fliess: 1887-1904*. Buenos Aires: Amorrortu editores.
- Freud, S. (1989a). Fragmentos da análise de um caso de histeria. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (3ª ed., Vol. 7). Rio de Janeiro: Imago.
- Freud, S. (1989b). O mal-estar na civilização. In *Edição standard brasileira das obras psicológicas completas de Sigmund Freud* (2ª ed., Vol. 21). Rio de Janeiro: Imago.
- Freud, S. (2001). *A interpretação dos sonhos*. Rio de Janeiro: Imago.
- Gamallo, P., & Garcia, M. (2013). FreeLing e TreeTagger: um estudo comparativo no âmbito do Português. *ProLNat Technical Report, 1*. Recuperado de [http://gramatica.usc.es/~gamallo/artigos-web/PROLNAT\\_Report\\_01.pdf](http://gramatica.usc.es/~gamallo/artigos-web/PROLNAT_Report_01.pdf)
- Garcia-Roza, L. A. (1994). A pesquisa de tipo teórico. *Psicanálise e universidade, 1*.
- Gaut, G., Steyvers, M., Imel, Z., Atkins, D., & Smyth, P. (2015).

- Content Coding of Psychotherapy Transcripts Using Labeled Topic Models. *IEEE Journal of Biomedical and Health Informatics*, PP(99), 1–1. doi: 10.1109/JBHI.2015.2503985
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (3ª ed.). Nova Iorque, Boca Raton, Londres: Chapman & Hall/CRC.
- Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. Recuperado 2015-03-11, de <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0521867061>
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology*. Cambridge: Cambridge University Press. Recuperado 2014-06-28, de <http://repository.cmu.edu/psychology/968/>
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America*, 101(Suppl 1), 5228–5235. Recuperado 2015-01-14, de <http://www.pnas.org/content/101/suppl.1/5228.long>
- Griffiths, T. L., Steyvers, M., Blei, D. M., & Tenenbaum, J. B. (2004). Integrating topics and syntax. In *Advances in neural information processing systems* (pp. 537–544). Recuperado 2015-03-14, de [http://machinelearning.wustl.edu/mlpapers/paper\\_files/NIPS2005\\_642.pdf](http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2005_642.pdf)
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114(2), 211–244. Recuperado 2014-06-24, de <http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.114.2.211> doi: 10.1037/0033-295X.114.2.211
- Grünbaum, A. (1977). Is Psychoanalysis a Pseudo-Science? Karl Popper versus Sigmund Freud. *Zeitschrift für philosophische Forschung*, 31(3), 333–353. Recuperado 2016-04-19, de <http://www.jstor.org/stable/20482837>
- Grünbaum, A. (1985). *The foundations of psychoanalysis: a philosophical critique*. Berkeley, Los Angeles, London: University of California Press.
- Guimarães, R. M., & Bento, V. E. S. (2008, Maio). O método do “estudo de caso” em psicanálise. *Psico*, 39(1). Recuperado 2015-04-14, de <http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/1484>

- Heinrich, G. (2009). *Parameter estimation for text analysis* (Rel. Téc.). Fraunhofer IGD: Technical report. Recuperado 2015-01-15, de <http://www.arbylon.net/publications/text-est2.pdf>
- Hoffmann, T. (1999). Probabilistic Latent Semantic Indexing. In *Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval*.
- Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015a). Computational psychotherapy research: Scaling up the evaluation of patient-provider interactions. *Psychotherapy*, 52(1), 19–30. Recuperado 2015-03-14, de <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0036841> doi: 10.1037/a0036841
- Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015b). Computational psychotherapy research: Scaling up the evaluation of patient-provider interactions. *Psychotherapy*, 52(1), 19–30. doi: 10.1037/a0036841
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An Introduction to Statistical Learning: with Applications in R*. Springer New York. Recuperado de <https://books.google.com.br/books?id=at1bmAEACAAJ>
- Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge, New York: Cambridge University Press.
- Jones, M., N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In *The Oxford handbook of computational and mathematical psychology*. New York: Oxford University Press.
- Jung, C. G. (1919). *Studies in word-association*. New York: Moffat, Year and Company.
- Jung, S. I., Serralta, F. B., Nunes, M. L. T., Eizirik, C. L., Jung, S. I., Serralta, F. B., ... Eizirik, C. L. (2014). Momentos distintos no abandono da psicoterapia psicanalítica. *Jornal Brasileiro de Psiquiatria*, 63(2), 133–141. Recuperado 2016-02-21, de [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S0047-20852014000200133&lng=pt&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0047-20852014000200133&lng=pt&nrm=iso&tlng=en) doi: 10.1590/0047-2085000000017
- Kächele, H., & Hölzer, M. (2007). Hartvig Dahl: The Lonely Rider. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 57(6), 233–234.
- Kächele, H., Schachter, J., & Thomä, H. (2011). *From Psychoanalytic Narrative to Empirical Single Case Research: Implications for Psychoanalytic Practice*. Taylor & Francis.
- Kruschke, J. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press. Recuperado 2015-03-11, de <http://>

- [www.indiana.edu/~kruschke/DoingBayesianDataAnalysis/](http://www.indiana.edu/~kruschke/DoingBayesianDataAnalysis/)
- Lakatos, I., Worrall, J., & Currie, G. (1980). *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers*. Cambridge University Press.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2-3), 259–284. Recuperado 2013-07-13, de <http://www.tandfonline.com/doi/abs/10.1080/01638539809545028>
- Laplanche, J. (1980). *Problématiques I: l'angoisse* (2ª ed.). Paris: Presses Universitaires de France.
- Laplanche, J. (2014). Interpreting (with) Freud. In *Seductions and enigmas: Laplanche, Theory and Culture* (pp. 171–184). Londres: Lawrence and Wishart. Recuperado 2015-03-14, de <http://www.palgrave-journals.com/pcs/journal/v11/n2/abs/2100077a.html>
- Laplanche, J., & Pontalis, J.-B. (1998). *Vocabulário de psicanálise* (3ª ed.). São Paulo: Martins Fontes.
- Leuzinger-Bohleber, E. E., & Kächele, H. (2015). *An open door review of outcome studies in psychoanalysis*. London: International psychoanalytical association.
- Li, W., & McCallum, A. (2008). Pachinko allocation: Scalable mixture models of topic correlations. *J. of Machine Learning Research*. Submitted. Recuperado 2015-03-14, de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.219.3333&rep=rep1&type=pdf>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22. Recuperado de <http://CRAN.R-project.org/doc/Rnews/>
- Luttinen, J. (2012). *BayesNet*. Recuperado de <https://github.com/jluttine/tikz-bayesnet>
- Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. Springer.
- Machado, J., & Cruz, S. M. S. C. d. S. (2011). Knowledge, reality and Physics teaching: modeling in a bungean inspiration. *Ciência & Educação (Bauru)*, 17(4), 887–902. Recuperado 2015-03-15, de [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S1516-73132011000400008&lng=en&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1516-73132011000400008&lng=en&nrm=iso&tlng=pt) doi: 10.1590/S1516-73132011000400008
- Mahony, P. (1990). *Freud, l'écrivain*. Paris: Les Belles Lettres.
- Meehl, P. (1983). Subjectivity in psychoanalytic inference: the nagging persistence of Wilhelm Fliess's Achensee question. In J. Earman

- (Ed.), *Minnesota studies in the philosophy of science* (Vol. 10, pp. 349–411). Minneapolis: University of Minnesota Press.
- Mezan, R. (1998). *Escrever a clínica* (2ª ed.). São Paulo: Casa do Psicólogo.
- Mezan, R. (2002). *Interfaces de psicanálise*. São Paulo: Companhia das Letras.
- Mezan, R. (2006, Junho). Pesquisa em psicanálise: algumas reflexões. *Jornal de Psicanálise*, 39(70), 227–241. Recuperado 2016-04-19, de [http://pepsic.bvsalud.org/scielo.php?script=sci\\_abstract&pid=S0103-58352006000100015&lng=pt&nrm=iso&tling=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S0103-58352006000100015&lng=pt&nrm=iso&tling=pt)
- Mezan, R. (2007). Que tipo de ciência é, afinal, a Psicanálise? *Natureza Humana*, 9(2), 319–359.
- Mimno, D., & Blei, D. (2011). Bayesian checking for topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 227–237). Association for Computational Linguistics. Recuperado 2015-02-25, de <http://dl.acm.org/citation.cfm?id=2145459>
- Papadimitriou, C. H., Raghavan, P., Tamaki, H., & Vempala, S. (1997). Latent semantic indexing: a probabilistic analysis. In *Proceedings of ACM PODS*. Tucson.
- Popper, K. (1962). *Conjectures and refutations: the growth of scientific knowledge*. New York, London: Basic Books.
- Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (n.d.). stm: R package for Structural Topic Models. *Journal of Statistical Software*.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2016). Navigating the local modes of big data: the case of topic models. In *Data analytics in social science, government, and industry*. New York: Cambridge University Press. Recuperado 2016-05-12, de <http://scholar.harvard.edu/files/dtingley/files/multimod.pdf>
- Roberts, M. E., Stewart, B. M., Tingley, D., & Airoldi, E. M. (2013). The Structural Topic Model and Applied Social Science. In *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*. Lake Tahoe, Nevada.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... Rand, D. G. (2014). Structural topic models



- for open-ended survey responses. *American Journal of Political Science*, 5(14).
- Sahlgren, M. (2006). *The Word-Space Model: using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces* (Tese de doutorado não publicada). Department of Linguistics, Stockholm University ; National Graduate School of Language Technology, Gothenburg University ; Swedish Institute of Computer Science Useware Laboratory, Stockholm; Göteborg; Kista.
- Schmid, H. (1995). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*. Recuperado de <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- Serralta, F. B. (2016, Agosto). Uncovering Interaction Structures in a Brief Psychodynamic Psychotherapy. *Paidéia (Ribeirão Preto)*, 26(64), 255–263. Recuperado 2016-02-13, de [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S0103-863X2016000200255&lng=en&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0103-863X2016000200255&lng=en&nrm=iso&tlng=en) doi: 10.1590/1982-43272664201613
- Serralta, F. B., Nunes, M. L. T., & Eizirik, C. L. (2007, Abril). Elaboração da versão em português do Psychotherapy Process Q-Set. *Revista de Psiquiatria do Rio Grande do Sul*, 29(1), 44–55. Recuperado 2016-01-15, de [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S0101-81082007000100011&lng=en&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0101-81082007000100011&lng=en&nrm=iso&tlng=pt) doi: 10.1590/S0101-81082007000100011
- Serralta, F. B., Nunes, M. L. T., & Eizirik, C. L. (2011, Dezembro). Methodological considerations concerning case studies in psychotherapy research. *Estudos de Psicologia (Campinas)*, 28(4), 501–510. Recuperado 2016-01-14, de [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S0103-166X2011000400010&lng=en&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0103-166X2011000400010&lng=en&nrm=iso&tlng=pt) doi: 10.1590/S0103-166X2011000400010
- Shachter, R. D. (2013, Janeiro). Bayes-Ball: The Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams). *arXiv:1301.7412 [cs]*. Recuperado 2016-08-12, de <http://arxiv.org/abs/1301.7412> (arXiv: 1301.7412)
- Silva, M. d. R., Barcellos, E. D. d., Sanchez, L. F., Steibel, D., Fernandes, P. d. P., Campezzatto, P. v. M., ... Klarmann,

- R. P. (2014, Dezembro). Registros de sessão terapêutica: relato, gravação ou transcrição? considerações sobre as diferenças entre os registros. *Psicologia Clínica*, 26(2), 121–138. Recuperado 2016-03-10, de [http://pepsic.bvsalud.org/scielo.php?script=sci\\_abstract&pid=S0103-56652014000200008&lng=pt&nrm=iso&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S0103-56652014000200008&lng=pt&nrm=iso&tlng=pt)
- Socher, R., Gershman, S., Sederberg, P., Norman, K., Perotte, A. J., & Blei, D. M. (2009). A Bayesian Analysis of Dynamics in Free Recall. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 1714–1722). Curran Associates, Inc. Recuperado 2015-01-08, de <http://papers.nips.cc/paper/3720-a-bayesian-analysis-of-dynamics-in-free-recall.pdf>
- Taddy, M. A. (2011, Setembro). On Estimation and Selection for Topic Models. *arXiv:1109.4518 [stat]*. Recuperado 2016-06-13, de <http://arxiv.org/abs/1109.4518> (arXiv: 1109.4518)
- Thomä, H., & Kächele, H. (1992). *Teoria e prática da psicanálise: fundamentos teóricos* (Vol. 1). Porto Alegre: Artes Médicas.
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1), 141–188. Recuperado 2013-07-09, de <http://www.aaai.org/Papers/JAIR/Vol137/JAIR-3705.pdf>
- Wallach, H. M., Minno, D., & McCallum, A. (2009). Rethinking LDA: Why priors matter. *NIPS*. Recuperado 2015-02-25, de [http://works.bepress.com/hanna\\_wallach/10/](http://works.bepress.com/hanna_wallach/10/)
- Wang, H., Zhang, D., & Zhai, C. (2011). Structural topic model for latent topical structure analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 1526–1535). Association for Computational Linguistics. Recuperado 2015-02-25, de <http://dl.acm.org/citation.cfm?id=2002657>
- Wang, X., & McCallum, A. (2006). Topics over time: a non-Markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 424–433). ACM. Recuperado 2015-02-23, de <http://dl.acm.org/citation.cfm?id=1150450>
- Wang, Y. (2008). *Distributed Gibbs Sampling of Latent Dirichlet Allocation: The Gritty Details*. Recuperado de <https://cxwangyi.files.wordpress.com/2012/01/llt.pdf>
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. Recuperado de <http://had.co.nz/>

ggplot2/book

- Wolpert, L., & Fonagy, P. (2009, Dezembro). There is no place for the psychoanalytic case report in the British Journal of Psychiatry. *The British Journal of Psychiatry*, 195(6), 483–487. Recuperado 2016-05-03, de <http://bjp.rcpsych.org/content/195/6/483> doi: 10.1192/bjp.bp.109.064451



# Apêndice A

## Código-fonte e visualização

O código-fonte das rotinas implementados em R para o tratamento e análise dos dados, bem como uma visualização interativa dos complexos do modelo final, pode ser conferidos no seguinte *link*:

<https://github.com/erikson84/TopicModelsFreeAssociation>